MODULATION OF REP PROTEIN ACTIVITY IN CLOSED-ENDED DNA (CEDNA) PRODUCTION

RELATED APPLICATIONS

[0001] This application claims priority to United States Provisional Application No. 62/806,076, filed on February 15, 2019, the contents of which is incorporated by reference in its entirety herein.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which is hereby incorporated by reference in its entirety. Said ASCII copy, created on February 13, 2020, is named 131698-05420_SL.txt and is 388,896 bytes in size.

TECHNICAL FIELD

[0003] The present invention relates to the field of gene therapy, including the delivery of exogenous DNA sequences to a target cell, tissue, organ or organism.

BACKGROUND

[0004] Gene therapy aims to improve clinical outcomes for patients suffering from either genetic mutations or acquired diseases caused by an aberration in the gene expression profile. Gene therapy includes the treatment or prevention of medical conditions resulting from defective genes or abnormal regulation or expression, *e.g.* underexpression or overexpression, that can result in a disorder, disease, malignancy, *etc*. For example, a disease or disorder caused by a defective gene might be treated, prevented or ameliorated by delivery of a corrective genetic material to a patient resulting in the therapeutic expression of the genetic material within the patient. The basis of gene therapy is to supply a transcription cassette with an active gene product (sometimes referred to as a transgene), *e.g.*, that can result in a positive gain-of-function effect, a negative loss-of-function effect, or another outcome, such as an oncolytic effect. Human monogenic disorders can be treated by the delivery and expression of a normal gene to the target cells. Delivery and expression of a corrective gene in the patient's target cells can be carried out *via* numerous methods, including the use of engineered viruses and viral gene delivery vectors. Among the many virus-derived vectors available (*e.g.,* recombinant retrovirus, recombinant lentivirus, recombinant adenovirus, and the like), recombinant adeno-associated virus (rAAV) is gaining popularity as a versatile vector in gene therapy.

[0005] Adeno-associated viruses (AAV) belong to the parvoviridae family and more specifically constitute the dependoparvovirus genus. The AAV genome is composed of a linear single-stranded DNA molecule which contains approximately 4.7 kilobases (kb) and consists of two major open reading frames (ORFs) encoding the non-structural Rep (replication) and structural Cap (capsid) proteins. A second ORF within the *cap* gene was identified that encodes the assembly-activating protein (AAP). The DNAs flanking the AAV

coding regions are two *cis*-acting inverted terminal repeat (ITR) sequences, approximately 145 nucleotides in length, with interrupted palindromic sequences that can be folded into energetically-stable hairpin structures that function as primers of DNA replication. In addition to their role in DNA replication, the ITR sequences have been shown to be involved in viral DNA integration into the cellular genome, rescue from the host genome or plasmid, and encapsidation of viral nucleic acid into mature virions (Muzyczka, (1992) Curr. Top. Micro. Immunol. 158:97-129).

[0006] Vectors derived from AAV (*i.e.,* recombinant AAV (rAVV) or AAV vectors) are attractive for delivering genetic material because (i) they are able to infect (transduce) a wide variety of non-dividing and dividing cell types including myocytes and neurons; (ii) they are devoid of the virus structural genes, thereby diminishing the host cell responses to virus infection, *e.g.*, interferon-mediated responses; (iii) wild-type viruses are considered non-pathologic in humans; (iv) in contrast to wild type AAV, which are capable of integrating into the host cell genome, replication-deficient AAV vectors lack the *rep* gene and generally persist as episomes, thus limiting the risk of insertional mutagenesis or genotoxicity; and (v) in comparison to other vector systems, AAV vectors are generally considered to be relatively poor immunogens and therefore do not trigger a significant immune response (see ii), thus gaining persistence of the vector DNA and potentially, long-term expression of the therapeutic transgenes. AAV vectors can also be produced and formulated at high titer and delivered *via* intra-arterial, intra-venous, or intra-peritoneal injections allowing vector distribution and gene transfer to significant muscle regions through a single injection in rodents (Goyenvalle *et al.*, 2004; Fougerousse *et al.*, 2007; Koppanati *et al.*, 2010; Wang *et al.*, 2009) and dogs. In a clinical study to treat spinal muscular dystrophy type 1, AAV vectors were delivered systemically with the intention of targeting the brain resulting in apparent clinical improvements.

[0007] However, there are several major deficiencies in using AAV particles as a gene delivery vector. One major drawback associated with rAAV is its limited viral packaging capacity of about 4.5 kb of heterologous DNA (Dong *et al.*, 1996; Athanasopoulos *et al.*, 2004; Lai *et al.*, 2010). As a result, use of AAV vectors has been limited to less than 150 kDa protein coding capacity. The second drawback is that as a result of the prevalence of wild-type AAV infection in the population, candidates for rAAV gene therapy have to be screened for the presence of neutralizing antibodies that eliminate the vector from the patient. A third drawback is related to the capsid immunogenicity that prevents re-administration to patients that were not excluded from an initial treatment. The immune system in the patient can respond to the vector which effectively acts as a "booster" shot to stimulate the immune system generating high titer anti-AAV antibodies that preclude future treatments. Some recent reports indicate concerns with immunogenicity in high dose situations. Another notable drawback is that the onset of AAV-mediated gene expression is relatively slow, given that single-stranded AAV DNA must be converted to double-stranded DNA prior to heterologous gene expression. While attempts have been made to circumvent this issue by constructing double-stranded DNA vectors, this strategy further limits the size of the transgene expression cassette that can be integrated into the AAV vector (McCarty, 2008; Varenika *et al.*, 2009; Foust *et al.*, 2009).

[0008] Additionally, conventional AAV virions with capsids are produced by introducing a plasmid or plasmids containing the AAV genome, rep genes, and cap genes (Grimm *et al.*, 1998). Upon introduction of these helper plasmids in *trans*, the AAV genome is "rescued" (*i.e.*, released and subsequently amplified) from the host genome, and is further encapsidated (viral capsids) to produce biologically active AAV vectors. However, such encapsidated AAV virus vectors were found to inefficiently transduce certain cell and tissue types. The capsids also induce an immune response.

[0009] Accordingly, use of adeno-associated virus (AAV) vectors for gene therapy is limited due to the single administration to patients (owing to the patient immune response), the limited range of transgene genetic material suitable for delivery in AAV vectors due to minimal viral packaging capacity (about 4.5kb) of the associated AAV capsid, as well as the slow AAV-mediated gene expression. The applications for rAAV clinical gene therapies are further encumbered by patient-to-patient variability not predicted by dose response in syngeneic mouse models or in other model species.

[0010] Recombinant capsid-free AAV vectors can be obtained as an isolated linear nucleic acid molecule comprising an expressible transgene and promoter regions flanked by two wild-type AAV inverted terminal repeat sequences (ITRs) including the Rep binding and terminal resolution sites (TRS). These recombinant AAV vectors are devoid of AAV capsid protein encoding sequences, and can be single-stranded, double-stranded or duplex with one or both ends covalently linked through the two wild-type ITR palindrome sequences (*e.g.*, WO2012/123430, U.S. Patent 9,598,703). They avoid many of the problems of AAV-mediated gene therapy in that the transgene capacity is much higher, transgene expression onset is rapid, and the patient immune system does recognize the DNA molecules as a virus to be cleared. However, constant expression of a transgene may not be desirable in all instances, and AAV canonical wild type ITRs may not be optimized for ceDNA function. Therefore, there remains an important unmet need for controllable recombinant DNA vectors as well as an improved production and/or expression properties.

SUMMARY

[0011] The invention described herein relates to an improved production of a non-viral capsid-free DNA vector with covalently-closed ends (referred to herein as a "closed-ended DNA vector" or a "ceDNA vector"). The ceDNA vectors produced by the methods as described herein are capsid-free, linear duplex DNA molecules formed from a continuous strand of complementary DNA with covalently-closed ends (linear, continuous and non-encapsidated structure), which comprise a 5' inverted terminal repeat (ITR) sequence and a 3' ITR sequence that are different, or asymmetrical with respect to each other.

[0012] The technology described herein relates to the production of a ceDNA vector or an AAV vector in a cell (*e.g.*, insect cell, mammalian cell) or in a cell free system with a *single Rep* protein species. In particular, the present disclosure is based, in part, on the surprising finding that either Rep78 or Rep68, alone, is sufficient for production of a ceDNA vector or an AAV vector in a cell. This is an improved and more efficient method of ceDNA vector production than described in the prior art,where AAV or ceDNA vectors are produced in cells (*e.g.*, insect cells) requiring *two* Rep proteins; for example, at least one small

Rep protein (*e.g.*, Rep52 or Rep40) and at least one large Rep protein (*e.g.*, Rep78 or Rep68). That is, the prior art describes that production of ceDNA vectors or AAV vectors is carried out using two Rep proteins, either encoded on separate nucleic acid constructs each operatively linked to a promoter, or two Rep proteins encoded on a single nucleic acid construct with two initiation sites, operatively linked to a single promoter.

[0013] Accordingly, one aspect of the technology described herein relates to a nucleic acid construct for the production of DNA vectors, *e.g.*, ceDNA vectors and other recombinant parvovirus (*e.g.* adeno-associated virus) vectors in cells (*e.g.* insect cells, mammalian cells) and cell free systems, where, for example, the insect cells or cell free system comprises a first nucleotide sequence encoding a single parvoviral Rep protein, where the nucleotide sequence does not have an open reading frame (ORF) and lacks a functional initiation codon downstream of the first initiation codon and/or lacks alternative splicing sites preventing exon skipping, thereby enabling the translation of only a single parvoviral Rep protein (*e.g.*, a Rep78 or Rep 68 protein) without the translation of additional Rep proteins at the later initiation codon (*e.g.*, any one or more of Rep52 or Rep 40) in the insect cells or cell free system. That is, a nucleic acid encoding Rep78 does not also produce a Rep52 protein, and similarly, a nucleic acid encoding Rep68 does not produce a Rep40 protein. Further no other Rep protein is present or expressed in the system.

[0014] In some embodiments, the methods and compositions described herein to use a single Rep protein can be used in the production of any ceDNA vector, including but not limited to, a ceDNA vector comprising asymmetric ITRS as disclosed in International Patent Application PCT/US18/49996, filed on September 7, 2018 (see, e.g, Examples 1-4); a ceDNA vector for gene editing as disclosed on the International Patent Application PCT/US18/64242 filed on December 6, 2018 (see, *e.g.*, Examples 1-7), or a ceDNA vector for production of antibodies or fusion proteins, as disclosed in the International Patent Application PCT/US19/18016, filed on February 14, 2019, (*e.g.*, see Examples 1-4), all of which are incorporated by reference in their entireties herein. In some embodiments, it is also envisioned that the methods and compositions described herein using a single Rep protein can be used in the synthetic production of a ceDNA vector, *e.g.*, in a cell free or insect-free system of ceDNA production, as disclosed in International Application PCT/US19/14122, filed on January 18, 2019, incorporated by reference in its entirety herein, where the single Rep protein can be used for protein-assisted ligation of the ITR oligonucleotides therein.

[0015] The technology described herein relates to an improved method of production of a ceDNA vector containing at least one modified AAV inverted terminal repeat sequence (ITR) and an expressible transgene. The ceDNA vectors disclosed herein can be produced according to the described methods in eukaryotic cells, thus devoid of prokaryotic DNA modifications and bacterial endotoxin contamination in insect cells.

[0016] Aspects of the invention relate to methods and compositions to produce ceDNA vectors and AAV vectors using a single Rep protein as described herein. Other embodiments relate to a ceDNA vector produced by the methods and compositions as provided herein.

[0017] In one aspect, non-viral capsid-free DNA vectors with covalently-closed ends produced by the methods as described herein are preferably linear duplex molecules, and are obtainable from a vector polynucleotide that encodes a heterologous nucleic acid operatively positioned between two different inverted terminal repeat sequences (ITRs) (*e.g.* AAV ITRs), wherein at least one of the ITRs comprises a terminal resolution site and a replication protein binding site (RPS) (sometimes referred to as a replicative protein binding site), *e.g.* a Rep binding site, and one of the ITRs comprises a deletion, insertion, or substitution with respect to the other ITR. That is, one of the ITRs is asymmetrical relative to the other ITR. In one embodiment, at least one of the ITRs is an AAV ITR, *e.g.* a wild type AAV ITR or modified AAV ITR. In one embodiment, at least one of the ITRs is a modified ITR relative to the other ITR – that is, the ceDNA comprises ITRs that are asymmetric relative to each other. In one embodiment, at least one of the ITRs is a non-functional ITR.

[0018] In some embodiments, a ceDNA vector produced by the methods and compositions as described herein comprises: (1) an expression cassette comprising a cis-regulatory element, a promoter and at least one transgene; or (2) a promoter operably linked to at least one transgene, and (3) two self-complementary sequences, *e.g.*, ITRs, flanking said expression cassette, wherein the ceDNA vector is not associated with a capsid protein. In some embodiments, the ceDNA vector comprises two self-complementary sequences found in an AAV genome, where at least one comprises an operative Rep-binding element (RBE) (also sometimes referred to herein as "RBS") and a terminal resolution site (trs) of AAV or a functional variant of the RBE, and one or more cis-regulatory elements operatively linked to a transgene. In some embodiments, the ceDNA vector comprises additional components to regulate expression of the transgene, for example, regulatory switches, which are described herein in the section entitled "Regulatory Switches" for controlling and regulating the expression of the transgene, and can include a regulatory switch, *e.g.*, a kill switch to enable controlled cell death of a cell comprising a ceDNA vector.

[0019] In some embodiments, the two self-complementary sequences can be ITR sequences from any known parvovirus, for example a dependovirus such as AAV (*e.g.*, AAV1-AAV12). Any AAV serotype can be used, including but not limited to a modified AAV2 ITR sequence, that retains a Rep-binding site (RBS) such as 5´-GCGCGCTCGCTCGCTC-3´ (SEQ ID NO: 531) and a terminal resolution site (trs) in addition to a variable palindromic sequence allowing for hairpin secondary structure formation. In some embodiments, the ITR is a synthetic ITR sequence that retains a functional Rep-binding site (RBS) such as 5'-GCGCGCTCGCTCGCTC-3' (SEQ ID NO: 531) and a terminal resolution site (TRS) in addition to a variable palindromic sequence allowing for hairpin secondary structure formation. In some examples, a modified ITR sequence retains the sequence of the RBS, trs and the structure and position of a Rep binding element forming the terminal loop portion of one of the ITR hairpin secondary structure from the corresponding sequence of the wild-type AAV2 ITR.

[0020] Exemplary ITR sequences for use in the ceDNA vectors produced by the methods and compositions as described herein can be any one or more of Tables 2-10A and 10B, or SEQ ID NO: 2, 52, 101-499 and

545-547 or the partial ITR sequences shown in FIG. 26A-26B. In some embodiments, the ceDNA vectors produced by the methods and compositions as described herein do not have an ITR that comprises any sequence selected from SEQ ID NOs: 500-529.

[0021] In some embodiments, a ceDNA vector produced by the methods and compositions as described herein can comprise an ITR with a modification in the ITR corresponding to any of the modifications in ITR sequences or ITR partial sequences shown in any one or more of Tables 2, 3, 4, 5, 6, 7, 8, 9, 10A and 10B herein.

[0022] As an exemplary example, a closed-ended DNA vector produced by the methods and compositions as described herein comprises a promoter operably linked to a transgene, where the ceDNA is devoid of capsid proteins and is: (a) produced from a ceDNA-plasmid (*e.g.*, see Examples 1-2 and/or FIGS. 1A-B) that encodes a mutated right side AAV2 ITR having the same number of intramolecularly duplexed base pairs as SEQ ID NO:2 or a mutated left side AAV2 ITR having the same number of intramolecularly duplexed base pairs as SEQ ID NO:51 in its hairpin secondary configuration (preferably excluding deletion of any AAA or TTT terminal loop in this configuration compared to these reference sequences), and (b) is identified as ceDNA using the assay for the identification of ceDNA by agarose gel electrophoresis under native gel and denaturing conditions in Example 1. Examples of such modified ITR sequences are provided in Tables 2, 3, 4, 5, 6, 7, 8, 9, 10A and 10B herein.

[0023] The technology described herein further relates to production of a ceDNA vector that can be used to deliver and encode one or more transgenes in a target cell, for example, where the ceDNA vector comprises a multicistronic sequence, or where the transgene and its native genomic context (*e.g.*, transgene, introns and endogenous untranslated regions) are together incorporated into the ceDNA vector. The transgenes can be protein encoding transcripts, non-coding transcripts, or both. The ceDNA vector produced by the methods and compositions as described herein can comprise multiple coding sequences, and a non-canonical translation initiation site or more than one promoter to express protein encoding transcripts, non-coding transcripts, or both. The transgene can comprise a sequence encoding more than one proteins, or can be a sequence of a non-coding transcript. The expression cassette can comprise, *e.g.*, more than 4000 nucleotides, 5000 nucleotides, 10,000 nucleotides or 20,000 nucleotides, or 30,000 nucleotides, or 40,000 nucleotides or 50,000 nucleotides, or any range between about 4000-10,000 nucleotides or 10,000-50,000 nucleotides, or more than 50,000 nucleotides. The ceDNA vectors produced by the methods and compositions as described herein do not have the size limitations of encapsidated AAV vectors, thus enable delivery of a large-size expression cassette to provide efficient expression of transgenes. In some embodiments, the ceDNA vector produced by the methods and compositions as described herein is devoid of prokaryote-specific methylation.

[0024] The expression cassette of a ceDNA vector produced by the methods and compositions as described herein can also comprise an internal ribosome entry site (IRES) and/or a 2A element. The cis-regulatory elements include, but are not limited to, a promoter, a riboswitch, an insulator, a mir-regulatable element, a

post-transcriptional regulatory element, a tissue- and cell type-specific promoter and an enhancer. In some embodiments the ITR can act as the promoter for the transgene. In some embodiments, the ceDNA vector comprises additional components to regulate expression of the transgene. For example, the additional regulatory component can be a regulator switch as disclosed herein, including but not limited to a kill switch, which can kill the ceDNA infected cell, if necessary, and other inducible and/or repressible elements.

[0025] The technology described herein further provides novel methods of efficiently producing a ceDNA vector or other AAV vector that can selectively express one or more transgenes. A ceDNA vector produced by the methods and compositions as described herein has the capacity to be taken up into host cells, as well as to be transported into the nucleus in the absence of the AAV capsid. In addition, the ceDNA vectors produced by the methods and compositions as described herein described herein lack a capsid and thus avoid the immune response that can arise in response to capsid-containing vectors.

[0026] In one embodiment, the capsid free non-viral DNA vector (ceDNA vector) is obtained from a plasmid (referred to herein as a "ceDNA-plasmid") comprising a polynucleotide expression construct template comprising in this order: a first 5' inverted terminal repeat (*e.g.* AAV ITR); an expression cassette; and a 3' ITR (*e.g.* AAV ITR), where at least one of the 5' and 3' ITR is a modified ITR, or where when both the 5' and 3' ITRs are modified, they have different modifications from one another and are not the same sequence. In such an embodiment, the ceDNA vector is obtained by the process as exemplified in the Examples and shown in FIG. 4A-4D herein, where only a single Rep protein is required for the production.

[0027] A ceDNA vector is obtainable by a number of means that would be known to the ordinarily skilled artisan after reading this disclosure. For example, a polynucleotide expression construct template used for generating the ceDNA vectors of the present invention can be a ceDNA-plasmid (*e.g.* see Table 12 or FIG. 10B), a ceDNA-bacmid, and/or a ceDNA-baculovirus. In one embodiment, the ceDNA-plasmid comprises a restriction cloning site (*e.g.* SEQ ID NO: 7) operably positioned between the ITRs where an expression cassette comprising *e.g.*, a promoter operatively linked to a transgene, *e.g.*, a reporter gene and/or a therapeutic gene) can be inserted. In some embodiments, ceDNA vectors are produced from a polynucleotide template (*e.g.*, ceDNA-plasmid, ceDNA-bacmid, ceDNA-baculovirus) containing an ITR modified as compared to the corresponding flanking AAV3 ITR or wild-type AAV2 ITR sequence, where the modification is any one or more of deletion, insertion, and/or substitution.

[0028] According to some aspects, the disclosure provides a method for producing a ceDNA vector in an insect cell (*e.g.*, Sf9, Sf21, Trichoplusia ni cells, and High Five cells) or mammalian cell (*e.g.*, HEK293, Huh-7, HeLa, HepG2, HeplA, 911, CHO, COS, MeWo, NIH3T3, A549, HT1080, monocytes, and mature and immature dendritic cells); the method comprising culturing an insect cell or mammalian cell comprising a first nucleotide sequence encoding a single parvoviral Rep protein, where the first nucleotide sequence lacks a functional initiation codon downstream of the first initiation codon and lacks alternative splicing sites preventing exon skipping, thereby enabling the translation of only a single Rep protein (*e.g.*, a Rep78)

without the translation of additional Rep proteins at the later initiation codon (*e.g.*, any one or more of Rep52 or Rep40) or a spliced variant of the full-length (*e.g.*, Rep68) in the cell.

[0029] According to some other aspects, the disclosure provides a method for producing a ceDNA vector in an insect cell (*e.g.*, Sf9, Sf21, Trichoplusia ni cells, and High Five cells) or mammalian cell (*e.g.*, HEK293, Huh-7, HeLa, HepG2, HeplA, 911, CHO, COS, MeWo, NIH3T3, A549, HT1080, monocytes, and mature and immature dendritic cells); the method comprising culturing an insect cell or mammalian cell comprising a first nucleotide sequence encoding a signle parvoviral Rep protein, wherein the first nucleotide sequence lacks a functional initiation codon downstream of the first inititation codon and contains a deletion of a carboxy terminal spliced sequence (*e.g.*, any portion or full-length of a c-terminal intron/skipped exon), thereby enabling the translation of only a signle Rep protein (*e.g.*, a Rep68) without the translation of additional Rep proteins at the later initiation codon (*e.g.*, any one or more of Rep52 or Rep40) or the full-length Rep72 protein in the cell.

[0030] According to some other aspects, the disclosure provides a method for producting a ceDNA vector in an insect cell (*e.g.*, Sf9, Sf21, Trichoplusia ni cells, High Five cells) or mammalian cell (*e.g.*, HEK293, Huh-7, HeLa, HepG2, HeplA, 911, CHO, COS, MeWo, NIH3T3, A549, HT1080, monocytes, and mature and immature dendritic cells); the method comprising culturing an insect cell or mammalian cell comprising a first nucleotide sequence encoding one or two Rep protein (*e.g.*, a Rep 78 and/or Rep68 protein), wherein the first nucleotide sequence lacks a functional initiation codon downstream of the first initiation codon and intact alternative splicing sites, thereby enabling the translation of a Rep78 and/or Rep68 protein only, without the translation of additional Rep proteins at the later initiation codon (*e.g.*, any one or more of Rep52 or Rep40).

[0031] The cell described in the methods above can further comprise a second nucleotide sequence comprising at least one AAV inverted terminal repeat (ITR) sequence flanking a heterologous sequence under conditions such that when the first sequence is expressed to produce Rep78 and/or Rep68, a ceDNA is produced by the Rep78 and/or Rep68 protein, without the presence of Rep52 or Rep40. The ceDNA vector then can be recovered from the cell. According to some embodiments, the nucleotide sequence comprising at least one AAV is part of an expression construct. According to some embodiments, the heterologous sequence comprises a therapeutic nucleic acid. According to some embodiments, the therapeutic nucleic acid is part of an expression construct. According to some embodiments, the cell further comprises a nucleic acid that serves as a marker. According to some embodiments, the nucleic acid that serves as a marker is part of an expression construct.

[0032] In a permissive host cell, in the presence of *e.g.*, a *single* Rep protein, the polynucleotide template having at least one modified ITR replicates to produce ceDNA vectors. ceDNA vector production undergoes two steps: (i) the single Rep proteins results in an excision ("rescue") step of template from the template backbone (*e.g.* ceDNA-plasmid, ceDNA-bacmid, ceDNA-baculovirus genome *etc.*), and (ii) the single Rep protein mediates replication of the excised ceDNA vector. The single Rep protein required for

the exision and replication steps (i) and (ii) can be any Rep protein described herein. Rep proteins and Rep binding sites of the various AAV serotypes are well known to those of ordinary skill in the art.

[0033] One of ordinary skill understands to choose a Rep protein from a serotype that binds to and replicates the nucleic acid sequence based upon at least one functional ITR. For example, if the replication competent ITR is from AAV serotype 2, the corresponding Rep would be from an AAV serotype that works with that serotype such as AAV2 ITR with AAV2 or AAV4 Rep but not AAV5 Rep, which does not. Upon replication (*i.e.*, after step (ii)), the covalently-closed ended ceDNA vector continues to accumulate in permissive cells and ceDNA vector is preferably sufficiently stable over time in the presence of the single Rep protein under standard replication conditions, *e.g.* to accumulate in an amount that is at least 1 pg/cell, preferably at least 2 pg/cell, preferably at least 3 pg/cell, more preferably at least 4 pg/cell, even more preferably at least 5 pg/cell.

[0034] Accordingly, one aspect of the invention relates to a process comprising the steps of: a) incubating a population of host cells (*e.g.* insect cells) harboring the polynucleotide expression construct template (*e.g.*, a ceDNA-plasmid, a ceDNA-bacmid, and/or a ceDNA-baculovirus), which is devoid of viral capsid coding sequences, in the presence of a *single* Rep protein under conditions effective and for a time sufficient to induce production of the ceDNA vector within the host cells, and wherein the host cells do not comprise viral capsid coding sequences; and b) harvesting and isolating the ceDNA vector from the host cells. The presence of a single Rep protein induces replication of the vector polynucleotide with a modified ITR to produce the ceDNA vector in a host cell. However, no viral particles (*e.g.* AAV virions) are expressed. Thus, there is no virion-enforced size limitation. It is envisioned that if the nucleic acid sequence encoding the Rep protein encodes a large Rep protein, *e.g.*, a Rep78 or Rep 68 protein, that the initiation codon for the smaller Rep proteins is modified such that only the large Rep protein is expressed in the cell.

[0035] According to some aspects, the disclosure provides an insect cell (*e.g.*, Sf9, Sf21, Trichoplusia ni cells, and High Five cells) or mammalian cell (*e.g.*, HEK293, Huh-7, HeLa, HepG2, HeplA, 911, CHO, COS, MeWo, NIH3T3, A549, HT1080, monocytes, and mature and immature dendritic cells); the insect cell or mammalian cell-line comprising a first nucleotide sequence encoding a single parvoviral Rep protein, where the first nucleotide sequence lacks a functional initiation codon downstream of the first initiation codon and lacks alternative splicing sites preventing exon skipping, thereby enabling the translation of only a single Rep protein (*e.g.*, a Rep78) without the translation of additional Rep proteins at the later initiation codon (*e.g.*, any one or more of Rep52 or Rep40) or a spliced variant of the full-length (*e.g.*, Rep68) in the cell.

[0036] According to some other aspects, the disclosure provides an insect cell (*e.g.*, Sf9, Sf21, Trichoplusia ni cells, and High Five cells) or mammalian cell (*e.g.*, HEK293, Huh-7, HeLa, HepG2, HeplA, 911, CHO, COS, MeWo, NIH3T3, A549, HT1080, monocytes, and mature and immature dendritic cells); the insect cell or mammalian cell comprising a first nucleotide sequence encoding a signle parvoviral Rep protein, wherein the first nucleotide sequence lacks a functional initiation codon downstream of the first

inititation codon and contains a deletion of a carboxy terminal spliced sequence (*e.g.*, any portion or full-length of a c-terminal intron/skipped exon), thereby enabling the translation of only a signle Rep protein (*e.g.*, a Rep68) without the translation of additional Rep proteins at the later initiation codon (*e.g.*, any one or more of Rep52 or Rep40) or the full-length Rep72 protein in the cell.

[0037] According to some other aspects, the disclosure provides an insect cell (*e.g.*, Sf9, Sf21, Trichoplusia ni cells, High Five cells) or mammalian cell (*e.g.*, HEK293, Huh-7, HeLa, HepG2, HeplA, 911, CHO, COS, MeWo, NIH3T3, A549, HT1080, monocytes, and mature and immature dendritic cells); the insect cell or mammalian cell-line comprising a first nucleotide sequence encoding one or two Rep protein (*e.g.*, a Rep 78 and/or Rep68 protein), wherein the first nucleotide sequence lacks a functional initiation codon downstream of the first initiation codon and intact alternative splicing sites, thereby enabling the translation of a Rep78 and/or Rep68 protein only, without the translation of additional Rep proteins at the later initiation codon (e.g., any one or more of Rep52 or Rep40)

[0038] The cell described above can further comprise a second nucleotide sequence comprising at least one AAV inverted terminal repeat (ITR) sequence flanking a heterologous sequence under conditions such that when the first sequence is expressed to produce Rep78 and/or Rep68, a ceDNA is produced by the Rep78 and/or Rep68 protein, without the presence of Rep52 or Rep40.The ceDNA vector then can be recovered from the cell. According to some embodiments, the nucleotide sequence comprising at least one AAV is part of an expression construct. According to some embodiments, the heterologous sequence comprises a therapeutic nucleic acid. According to some embodiments, the therapeutic nucleic acid is part of an expression construct. According to some embodiments, the cell further comprises a nucleic acid that serves as a marker. According to some embodiments, the nucleic acid that serves as a marker is part of an expression construct.

[0039] According to some aspects, the disclosure provides a cell free system comprising a first nucleotide sequence encoding a single parvoviral Rep protein, where the nucleotide sequence lacks a functional initiation codon downstream of the first initiation codon and/or lacks alternative splicing sites preventing exon skipping, thereby enabling the translation of only a single parvoviral Rep protein (*e.g.*, a Rep78 or Rep 68 protein) without the translation of additional Rep proteins at the later initiation codon (*e.g.*, any one or more of Rep52 or Rep 40) in the cell free system. According to some embodiments, a nucleic acid encoding Rep78 does not also produce a Rep52 or Rep40 protein. According to some embodiments, a nucleic acid encoding Rep68 does not produce a Rep52 or Rep40 protein. According to some embodiments, the insect cell, the mammalian cell or the cell free system does not express any other Rep protein.

[0040] A ceDNA vector produced according to the methods as described herein using a single Rep protein, is isolated from the host cells, and its presence can be confirmed by digesting DNA isolated from the host cell with a restriction enzyme having a single recognition site on the ceDNA vector and analyzing the digested DNA material on denaturing and non-denaturing gels to confirm the presence of characteristic bands of linear and continuous DNA as compared to linear and non-continuous DNA.

Description of Drawings

[0041] FIG. 1A illustrates an exemplary structure of a ceDNA vector produced using a single Rep protein according to the methods and compositions as described herein. In this embodiment, the exemplary ceDNA vector comprises an expression cassette containing CAG promoter, WPRE, and BGHpA. An open reading frame (ORF) encoding a transgene is inserted into the cloning site (R3/R4) between the CAG promoter and WPRE. The expression cassette is flanked by two inverted terminal repeats (ITRs) – the wild-type AAV2 ITR on the upstream (5'-end) and the modified ITR on the downstream (3'-end) of the expression cassette, therefore the two ITRs flanking the expression cassette are asymmetric with respect to each other. A person of ordinary skill in the art will appreciate that any ITR can be used. For exemplary purposes, the ITRs in the ceDNA constructs in this Figure and in the Examples herein are a modified ITR and a WT ITR. However, encompassed herein are ceDNA vectors that contain a heterologous nucleic acid sequence (*e.g.*, a transgene) positioned between any two inverted terminal repeat (ITR) sequences, where the ITR sequences can be an asymmetrical ITR pair or a symmetrical- or substantially symmetrical ITR pair, as these terms are defined herein. A ceDNA vector as disclosed herein can comprise ITR sequences that are selected from any of: (i) at least one WT ITR and at least one modified AAV inverted terminal repeat (mod-ITR) (*e.g.*, asymmetric modified ITRs); (ii) two modified ITRs where the mod-ITR pair have a different three-dimensional spatial organization with respect to each other (*e.g.*, asymmetric modified ITRs), or (iii) symmetrical or substantially symmetrical WT-WT ITR pair, where each WT-ITR has the same three-dimensional spatial organization, or (iv) symmetrical or substantially symmetrical modified ITR pair, where each mod-ITR has the same three-dimensional spatial organization. In some embodiments, the methods of the present disclosure encompass using a single rep protein for production of a ceDNA vector that is formulated in a composition that includes a delivery system, such as but not limited to a liposome nanoparticle delivery system.

[0042] FIG. 1B illustrates an exemplary structure of a ceDNA vector produced using a single Rep protein according to the methods and compositions as described herein, where the ceDNA vector comprises an expression cassette containing CAG promoter, WPRE, and BGHpA. An open reading frame (ORF) encoding Luciferase transgene is inserted into the cloning site between CAG promoter and WPRE. The expression cassette is flanked by two inverted terminal repeats (ITRs) – a modified ITR on the upstream (5'-end) and a wild-type ITR on the downstream (3'-end) of the expression cassette. As discussed in **FIG. 1A**, a skilled artisan can readily select ITR sequences to be an asymmetrical ITR pair or a symmetrical- or substantially symmetrical ITR pair, as these terms are defined herein.

[0043] FIG. 1C illustrates an exemplary structure of a ceDNA vector produced using a single Rep protein according to the methods and compositions as described herein, where the ceDNA vector comprises an expression cassette containing an enhancer/promoter, an open reading frame (ORF) for insertion of a transgene, a post transcriptional element (WPRE), and a polyA signal. An open reading frame (ORF) allows insertion of a transgene into the cloning site between CAG promoter and WPRE. The expression cassette is

flanked by two inverted terminal repeats (ITRs) that are asymmetrical with respect to each other; a modified ITR on the upstream (5'-end) and a modified ITR on the downstream (3'-end) of the expression cassette, where the 5' ITR and the 3'ITR are both modified ITRs but have different modifications (*i.e.*, they do not have the same modifications). As discussed in **FIG. 1A**, a skilled artisan can readily select ITR sequences to be an asymmetrical ITR pair or a symmetrical- or substantially symmetrical ITR pair, as these terms are defined herein.

[0044] FIG. 2A provides the T-shaped stem-loop structure of a wild-type left ITR of AAV2 (SEQ ID NO: 538) with identification of A-A' arm, B-B' arm, C-C' arm, two Rep binding sites (RBE and RBE') and also shows the terminal resolution site (*trs*). The RBE contains a series of 4 duplex tetramers that are believed to interact with either Rep78 or Rep68. In addition, the RBE' is also believed to interact with Rep complex assembled on the wild-type ITR or mutated ITR in the construct. The D and D' regions contain transcription factor binding sites and other conserved structure. **FIG. 2B** shows proposed Rep-catalyzed nicking and ligating activities in a wild-type left ITR (SEQ ID NO: 539), including the T-shaped stem-loop structure of the wild-type left ITR of AAV2 with identification of A-A' arm, B-B' arm, C-C' arm, two Rep Binding sites (RBE and RBE') and also shows the terminal resolution site (*TRS*), and the D and D' region comprising several transcription factor binding sites and other conserved structure.

[0045] FIG. 3A provides the primary structure (polynucleotide sequence) (left) and the secondary structure (right) of the RBE-containing portions of the A-A' arm, and the C-C' and B-B' arm of the wild type left AAV2 ITR (SEQ ID NO: 540). **FIG. 3B** shows an exemplary mutated ITR (also referred to as a modified ITR) sequence for the left ITR. Shown is the primary structure (left) and the predicted secondary structure (right) of the RBE portion of the A-A' arm, the C arm and B-B' arm of an exemplary mutated left ITR (ITR-1, left) (SEQ ID NO: 113). **FIG. 3C** shows the primary structure (left) and the secondary structure (right) of the RBE-containing portion of the A-A' loop, and the B-B' and C-C' arms of wild type right AAV2 ITR (SEQ ID NO: 541). **FIG. 3D** shows an exemplary right modified ITR. Shown is the primary structure (left) and the predicted secondary structure (right) of the RBE containing portion of the A-A' arm, the B-B' and the C arm of an exemplary mutant right ITR (ITR-1, right) (SEQ ID NO: 114). Any combination of left and right ITR (*e.g.*, AAV2 ITRs or other viral serotype or synthetic ITRs) can be used, provided the left ITR is asymmetric or different from the right ITR. Each of FIGS. 3A-3D polynucleotide sequences refer to the sequence used in the plasmid or bacmid/baculovirus genome used to produce the ceDNA as described herein. Also included in each of **FIGS. 3A-3D** are corresponding ceDNA secondary structures inferred from the ceDNA vector configurations in the plasmid or bacmid/baculovirus genome and the predicted Gibbs free energy values.

[0046] FIG. 4A is a schematic illustrating an upstream process for making baculovirus infected insect cells (BIICs) that are useful in the production of ceDNA in the process described in the schematic in FIG. 4B. In this embodiments, two bacmids are generated by transposing a ceDNA plasmid or Rep-plasmid (encoding a single Rep protein) into a baculovirus expression vector to generate a ceDNA vector bacmid (*i.e.*, Bacmid-

1) and a single Rep Bacmid (Rep-Bacmid), which are used to transfect insect cells to produce baculovirus injected insect cells, BIIC-1 and BICC-2 (single Rep), respectively. **FIG. 4B** is a schematic of an exemplary method of ceDNA production using the insect cells (*e.g.*, BICC-2) comprising the Rep-Bacmid comprising the nucleic acid sequence for a single Rep protein, and **FIG. 4C** illustrates a biochemical method and process to confirm ceDNA vector production using the single Rep protein methodology described herein. **FIG. 4D** and **FIG. 4E** are schematic illustrations describing a process for identifying the presence of ceDNA in DNA harvested from cell pellets obtained during the ceDNA production processes in **FIG. 4B**. **FIG. 4E** shows DNA having a non-continuous structure. The ceDNA can be cut by a restriction endonuclease, having a single recognition site on the ceDNA vector, and generate two DNA fragments with different sizes (1kb and 2kb) in both neutral and denaturing conditions. **FIG. 4E** also shows a ceDNA having a linear and continuous structure. The ceDNA vector can be cut by the restriction endonuclease, and generate two DNA fragments that migrate as 1kb and 2kb in neutral conditions, but in denaturing conditions, the stands remain connected and produce single strands that migrate as 2kb and 4kb. **FIG. 4D** shows schematic expected bands for an exemplary ceDNA either left uncut or digested with a restriction endonuclease and then subjected to electrophoresis on either a native gel or a denaturing gel. The leftmost schematic is a native gel, and shows multiple bands suggesting that in its duplex and uncut form ceDNA exists in at least monomeric and dimeric states, visible as a faster-migrating smaller monomer and a slower-migrating dimer that is twice the size of the monomer. The schematic second from the left shows that when ceDNA is cut with a restriction endonuclease, the original bands are gone and faster-migrating (*e.g.*, smaller) bands appear, corresponding to the expected fragment sizes remaining after the cleavage. Under denaturing conditions, the original duplex DNA is single-stranded and migrates as a species twice as large as observed on native gel because the complementary strands are covalently linked. Thus, in the second schematic from the right, the digested ceDNA shows a similar banding distribution to that observed on native gel, but the bands migrate as fragments twice the size of their native gel counterparts. The rightmost schematic shows that uncut ceDNA under denaturing conditions migrates as a single-stranded open circle, and thus the observed bands are twice the size of those observed under native conditions where the circle is not open. In this figure "kb" is used to indicate relative size of nucleotide molecules based, depending on context, on either nucleotide chain length (*e.g.*, for the single stranded molecules observed in denaturing conditions) or number of basepairs (*e.g.*, for the double-stranded molecules observed in native conditions).

[0047] FIG. 5 is an exemplary picture of a denaturing gel running examples of ceDNA vectors with (+) or without (-) digestion with endonucleases (EcoRI for ceDNA construct 1 and 2; BamH1 for ceDNA construct 3 and 4; SpeI for ceDNA construct 5 and 6; and XhoI for ceDNA construct 7 and 8). Sizes of bands highlighted with an asterisk were determined and provided on the bottom of the picture.

[0048] FIG. 6A shows results from an *in vitro* protein expression assay measuring Luciferase activity (y-axis, RQ (Luc)) in HEK293 cells 48 hours after transfection of 400 ng (black), 200 ng (gray), or 100 ng (white) of the constructs identified on the x-axis (construct-1, construct-3, construct-5, construct-7 (Table 12). **FIG. 6B** shows Luciferase activity (y-axis, RQ (Luc)) measured in HEK293 cells 48 hours after

transfection of 400 ng (black), 200 ng (gray), or 100 ng (white) of the constructs identified on the x-axis (construct -2, construct -4, construct -6, construct-8) (Table 12). Luciferase activities measured in HEK293 cells treated with Fugene without any plasmids ("Fugene"), or in untreated HEK293 cells ("Untreated") are also provided.

[0049] FIG. 7A shows viability of HEK293 cells (y-axis) 48 hours after transfection of 400 ng (black), 200 ng (gray), or 100 ng (white) of the constructs identified on the x-axis (construct-1, construct-3, construct-5, construct-7). **FIG. 7B** shows viability of HEK293 cells (y-axis) 48 hours after transfection of 400 ng (black), 200 ng (gray), or 100 ng (white) of the constructs identified on the x-axis (construct-2, construct-4, construct-6, construct-8).

[0050] FIG. 8A is an exemplary Rep-bacmid in the pFBDLSR plasmid comprising the nucleic acid sequences for modified Rep78 protein, where the modified Rep 78 protein is modification of amino acid residue 225 (Met) of SEQ ID NO: 530, wherein the amino acid residue 225 is changed to a glycine (Gly) (*e.g.*, M225G or Met225Gly) or threonine (Thr) (*e.g.*, M225T or Met225Thr). This exemplary Rep-bacmid comprises: IE1 promoter fragment (SEQ ID NO:66); Rep78 nucleotide sequence encoding a modified Rep78 protein that lacks a functional initiation codon downstream of the first initiation codon, thereby enabling translation of a single Rep78 protein. As one of skilled in the art will appreciate, one can modify this modified Rep78 bacmid or modified Rep78 plasmid with the nucleic acid encoding any single Rep protein (*e.g.*, Rep68, Rep52, Rep40) that has been modified to have a single initiation codon and therefore encodes a single Rep protein. **FIG. 8B** is a schematic of an exemplary ceDNA-plasmid-1, with the wt-L ITR, CAG promoter, luciferase transgene, WPRE and polyadenylation sequence, and mod-R ITR.

[0051] FIG. 9A shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the C-C' arm of an exemplary modified left ITR ("ITR-2 (Left)" SEQ ID NO: 101) and **FIG. 9B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm of an exemplary a modified right ITR ("ITR-2 (Right)" SEQ ID NO: 102). They are predicted to form a structure with a single arm (C-C') and a single unpaired loop. Their Gibbs free energies of unfolding are predicted to be -72.6 kcal/mol.

[0052] FIG. 10A shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the B-B' arm of an exemplary modified left ITR ("ITR-3 (Left)" SEQ ID NO: 103) and **FIG. 10B** shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the B-B' arm of an exemplary modified right ITR ("ITR-3 (Right)" SEQ ID NO: 104). They are predicted to form a structure with a single arm (B-B') and a single unpaired loop. Their Gibbs free energies of unfolding are predicted to be -74.8 kcal/mol.

[0053] FIG. 11A shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the C-C' arm of an exemplary modified left ITR ("ITR-4 (Left)" SEQ ID NO: 105) and **FIG. 11B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm of an exemplary modified right ITR ("ITR-4 (Right)" SEQ ID NO: 106). They are predicted to form a

structure with a single arm (C-C') and a single unpaired loop. Their Gibbs free energies of unfolding are predicted to be -76.9 kcal/mol.

[0054] FIG. 12A shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the C-C' and B-B' portions of an exemplary modified left ITR, showing complementary base pairing of the C-B' and C'-B portions ("ITR-10 (Left)" SEQ ID NO: 107) and **FIG. 12B** shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the B-B' and C-C' portions of an exemplary modified right ITR, showing complementary base pairing of the B-C' and B'-C portions ("ITR-10 (Right)" SEQ ID NO: 108). They are predicted to form a structure with a single arm (a portion of C'-B and C-B' or a portion of B'-C and B-C') and a single unpaired loop. Their Gibbs free energies of unfolding are predicted to be -83.7 kcal/mol.

[0055] FIG. 13A shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the C-C' and B-B' portions of an exemplary modified left ITR ("ITR-17 (Left)" SEQ ID NO: 109) and **FIG. 13B** shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the C-C' and B-B' portions of an exemplary modified right ITR ("ITR-17 (Right)" SEQ ID NO: 110). Both ITR-17 (left) and ITR-17 (right) are predicted to form a structure with a single arm (B-B') and a single unpaired loop. Their Gibbs free energies of unfolding are predicted to be -73.3 kcal/mol.

[0056] FIG. 14A shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm of an exemplary modified ITR ("ITR-6 (Left)" SEQ ID NO: 111) and **FIG. 14B** shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm of an exemplary modified ITR ("ITR-6 (Right)" SEQ ID NO: 112). Both ITR-6 (left) and ITR-6 (right) are predicted to form a structure with a single arm. Their Gibbs free energies of unfolding are predicted to be -54.4 kcal/mol.

[0057] FIG. 15A shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the C arm and B-B' arm of an exemplary a modified left ITR ("ITR-1 (Left)" SEQ ID NO: 113) and **FIG. 15B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C arm and B-B' arm of an exemplary modified right ITR ("ITR-1 (Right)" SEQ ID NO: 114). Both ITR-1 (left) and ITR-1 (right) are predicted to form a structure with two arms, one of which is truncated. Their Gibbs free energies of unfolding are predicted to be -74.7 kcal/mol.

[0058] FIG. 16A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C' arm and B-B' arm of an exemplary modified left ITR ("ITR-5 (Left)" SEQ ID NO: 545) and **FIG. 16B** shows the predicted lowest energy structure of the RBE containing portion of the A-A' arm and the B-B' arm and C' arm of an exemplary modified right ITR ("ITR-5 (Right)" SEQ ID NO: 116). Both ITR-5 (left) and ITR-5 (right) are predicted to form a structure with two arms, one of which is (*e.g.*, the C' arm) truncated. Their Gibbs free energies of unfolding are predicted to be -73.4 kcal/mol.

[0059] FIG. 17A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-B' arm of an exemplary modified left ITR ("ITR-7 (Left)" SEQ ID NO: 117) and **FIG. 17B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm

and the B-B' arm and C-C' arm of an exemplary modified right ITR ("ITR-7 (Right)" SEQ ID NO: 118). Both ITR-17 (left) and ITR-17 (right) are predicted to form a structure with two arms, one of which (*e.g.*, B-B' arm) is truncated. Their Gibbs free energies of unfolding are predicted to be -89.6 kcal/mol.

[0060] FIG. 18A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-B' arm of an exemplary modified left ITR ("ITR-8 (Left)" SEQ ID NO: 119) and **FIG. 18B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the B-B' arm and C-C' arm of an exemplary modified right ITR ("ITR-8 (Right)" SEQ ID NO: 120). Both ITR-8 (left) and ITR-8 (right) are predicted to form a structure with two arms, one of which is truncated. Their Gibbs free energies of unfolding are predicted to be -86.9 kcal/mol.

[0061] FIG. 19A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-B' arm of an exemplary modified left ITR ("ITR-9 (Left)" SEQ ID NO: 121) and **FIG. 19B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the B-B' arm and C-C' arm of an exemplary modified right ITR ("ITR-9 (Right)" SEQ ID NO: 122). Both ITR-9 (left) and ITR-9 (right) are predicted to form a structure with two arms, one of which is truncated. Their Gibbs free energies of unfolding are predicted to be -85.0 kcal/mol.

[0062] FIG. 20A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-B' arm of an exemplary modified left ITR ("ITR-11 (Left)" SEQ ID NO: 123) and **FIG. 20B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the B-B' arm and C-C' arm of an exemplary modified right ITR ("ITR-11 (Right)" SEQ ID NO: 124). Both ITR-11 (left) and ITR-11 (right) are predicted to form a structure with two arms, one of which is truncated. Their Gibbs free energies of unfolding are predicted to be -89.5 kcal/mol.

[0063] FIG. 21A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-B' arm of an exemplary modified left ITR ("ITR-12 (Left)" SEQ ID NO: 125) and **FIG. 21B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the B-B' arm and C-C' arm of an exemplary modified right ITR ("ITR-12 (Right)" SEQ ID NO: 126). Both ITR-12 (left) and ITR-12 (right) They are predicted to form a structure with two arms, one of which is truncated. Their Gibbs free energies of unfolding are predicted to be -86.2 kcal/mol.

[0064] FIG. 22A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-B' arm of an exemplary modified left ITR ("ITR-13 (Left)" SEQ ID NO: 127) and **FIG. 22B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the B-B' arm and C-C' arm of an exemplary a modified right ITR ("ITR-13 (Right)" SEQ ID NO: 128). Both ITR-13 (left) and ITR-13 (right) are predicted to form a structure with two arms, one of which (*e.g.*, C-C' arm) is truncated. Their Gibbs free energies of unfolding are predicted to be -82.9 kcal/mol.

[0065] FIG. 23A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-B' arm of an exemplary modified left ITR ("ITR-14 (Left)" SEQ ID NO: 129) and **FIG. 23B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm

and the B-B' arm and C-C' arm of an exemplary modified right ITR ("ITR-14 (Right)" SEQ ID NO: 130). Both ITR-14 (left) and ITR-14 (right) are predicted to form a structure with two arms, one of which (*e.g.*, C-C' arm) is truncated. Their Gibbs free energies of unfolding are predicted to be -80.5 kcal/mol.

[0066] FIG. 24A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-C' arm of an exemplary modified left ITR ("ITR-15 (Left)" SEQ ID NO: 131) and **FIG. 24B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the B-B' arm and C-C' arm of an exemplary modified right ITR ("ITR-15 (Right)" SEQ ID NO: 132). Both ITR-15 (left) and ITR-15 (right) are predicted to form a structure with two arms, one of which (*e.g.*, the C-C' arm) is truncated. Their Gibbs free energies of unfolding are predicted to be -77.2 kcal/mol.

[0067] FIG. 25A shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the C-C' arm and B-C' arm of an exemplary modified left ITR ("ITR-16 (Left) SEQ ID NO: 133) and **FIG. 25B** shows the predicted lowest energy structure of the RBE-containing portion of the A-A' arm and the B-B' arm and C-C' arm of an exemplary a modified right ITR ("ITR-16 (Right)" SEQ ID NO: 134). Both ITR-16 (left) and ITR-16 (right) are predicted to form a structure with two arms, one of which (*e.g.*, C-C' arm) is truncated. Their Gibbs free energies of unfolding are predicted to be -73.9 kcal/mol.

[0068] FIG. 26A shows predicted structures of the RBE-containing portion of the A-A' arm and modified B-B' arm and/or modified C-C' arm of exemplary modified right ITRs listed in Table 10A. **FIG. 26B** shows predicted structures of the RBE-containing portion of the A-A' arm and modified C-C' arm and/or modified B-B' arm of exemplary modified left ITRs listed in Table 10B. The structures shown are the predicted lowest free energy structure. Color code: red = >99% probability; orange = 99%-95% probability; beige = 95-90% probability; dark green 90%-80%; bright green = 80%-70%; light blue = 70%-60%; dark blue 60%-50% and pink = < 50%.

[0069] FIG. 27 shows luciferase activity of Sf9 GlycoBac insect cells transfected with selected asymmetric ITR mutant variants from Table 10A and 10B. The ceDNA vector had a luciferase gene flanked by a wt ITR and a modified asymmetric ITR selected from Table 10A or 10B. "ITR-50 R no rep" is the known rescuable mutant without co-infection of Rep containing baculovirus. "Mock" conditions are transfection reagents only, without donor DNA.

[0070] FIG. 28 shows a native agarose gel (1% agarose, 1x TAE buffer) of representative crude ceDNA extracts from Sf9 insect cell cultures transfected with ceDNA-plasmids comprising a Left wt-ITR with the other ITR selected from various mutant Right ITRs disclosed in Table 10A. 2ug of total extract was loaded per lane. From left to right: Lane 1) 1 kb plus ladder, Lane 2) ITR-18 Right, Lane 3) ITR-49 Right Lane 4) ITR-19 Right, Lane 5) ITR-20 Right, Lane 6) ITR-21 Right, Lane 7) ITR-22 Right, Lane 8) ITR-23 Right, Lane 9) ITR-24 Right, Lane 10) ITR-25 Right, Lane 11) ITR-26 Right, Lane 12) ITR-27 Right, Lane 13) ITR-28 Right, Lane 14) ITR-50 Right, lane 15) 1 kb plus ladder.

[0071] FIG. 29 shows a denaturing gel (0.8 % alkaline agarose) of representative constructs from ITR mutant library. The ceDNA vector is produced from plasmids constructs comprising a Left wt-ITR with the

other ITR selected from various mutant Right ITRs disclosed in Table 10A. From left to right, Lane 1) 1kb Plus DNA Ladder, Lane 2) ITR-18 Right un-cut, Lane 3) ITR-18 Right restriction digest, Lane 4) ITR-19 Right un-cut, Lane 5) ITR-19 Right restriction digest, Lane 6) ITR-21 Right un-cut, Lane 7) ITR-21 Right restriction digest, Lane 8) ITR-25 Right un-cut, Lane 9) ITR-25 Right restriction digest. Extracts were treated with EcoRI restriction endonuclease. Each mutant ceDNA is expected to have a single EcoRI recognition site, producing two characteristic fragments, ~2,000 bp and ~3,000 bp, which will run at ~4,000 and ~6,000 bp, respectively, under denaturing conditions. Untreated ceDNA extracts are ~5,000 bp and expected to migrate at ~11,000 bp under denaturing conditions.

[0072] FIG. 30 shows luciferase activity *in vitro* in HEK293 cells of ITR mutants ITR-18 Right, ITR-19 Right, ITR-21 Right and ITR-25 Right, and ITR-49**,** where the left ITR in the ceDNA vector is WT ITR. "Mock" conditions are transfection reagents only, without donor DNA, and untreated is the negative control.

[0073] FIG. 31 is a table showing various properties and activities (*e.g.*, DNA binding, DNA nicking, helicase activity, ATPase activity and Zn finger activities) of different Rep protein species (*e.g.*, wild-type Rep78, wild type Rep68, wild type Rep52 and wild type Rep40) and modified Rep68 species, *e.g.*, where the amino acid of Rep78 protein is modified to any of Y156, K340H, Met → Gly (M225G). The modification of Rep78 of Met→ Gly (M225G) maintained all properties and activities of the wild-type Rep78 protein.

[0074] FIGS. 32A and 32B are non-denaturing gels showing the presence of the highly stable DNA vectors and characteristic bands confirming the presence of the highly stable close-ended DNA (ceDNA) vector made with a single Rep protein using methods described herein. In **Fig. 32A**, higher amounts of ceDNA vector are produced using a nucleic acid of modified Rep78 with the modification of Rep78 of Met→ Gly (M225G) (lane 1) or Rep Met→Thr (M225T) (lane 2) as compared to the production using nucleic acid encoding wild-type Rep78 (lane 5) where the nucleic acid expresses both the Rep78/68 protein and the Rep52/40 protein. No ceDNA vector was produced with Rep78 binding/nicking mutants, comprising modifications of Gly (Y156F) (lane 3) or Thr (Y156F) (lane 4). In **Fig. 32B**, Rep68 Met→ Gly (M225G) and Rep68 Met→Thr (M225T) mutants also produced ceDNA vector, to levels equal to or greater than amounts of ceDNA vector produced using a nucleic acid of modified Rep78 with the modification of Rep78 of Met→ Gly (M225G) or Rep Met→Thr (M225T). DLSR: a plasmid construct expressing long (Rep78) and short (Rep52) Rep protein in tandem; pIE78: wildtype full-length Rep78 sequence; Rep78 M→G: full length Rep78 containing M225G single mutation; Rep78M→T: full length Rep78 containing M225T single mutation; Rep78Y156F: full length Rep78 having a single mutation in nickase domain.

DETAILED DESCRIPTION

I. Definitions

[0075] Unless otherwise defined herein, scientific and technical terms used in connection with the present application shall have the meanings that are commonly understood by those of ordinary skill in the art to

which this disclosure belongs. It should be understood that this invention is not limited to the particular methodology, protocols, and reagents, etc., described herein and as such can vary. The terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which is defined solely by the claims. Definitions of common terms in immunology and molecular biology can be found in The Merck Manual of Diagnosis and Therapy, 19th Edition, published by Merck Sharp & Dohme Corp., 2011 (ISBN 978-0-911910-19-3); Robert S. Porter *et al.* (eds.), Fields Virology, 6th Edition, published by Lippincott Williams & Wilkins, Philadelphia, PA, USA (2013), Knipe, D.M. and Howley, P.M. (ed.), The Encyclopedia of Molecular Cell Biology and Molecular Medicine, published by Blackwell Science Ltd., 1999-2012 (ISBN 9783527600908); and Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8); Immunology by Werner Luttmann, published by Elsevier, 2006; Janeway's Immunobiology, Kenneth Murphy, Allan Mowat, Casey Weaver (eds.), Taylor & Francis Limited, 2014 (ISBN 0815345305, 9780815345305); Lewin's Genes XI, published by Jones & Bartlett Publishers, 2014 (ISBN-1449659055); Michael Richard Green and Joseph Sambrook, Molecular Cloning: A Laboratory Manual, 4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., USA (2012) (ISBN 1936113414); Davis *et al.*, Basic Methods in Molecular Biology, Elsevier Science Publishing, Inc., New York, USA (2012) (ISBN 044460149X); Laboratory Methods in Enzymology: DNA, Jon Lorsch (ed.) Elsevier, 2013 (ISBN 0124199542); Current Protocols in Molecular Biology (CPMB), Frederick M. Ausubel (ed.), John Wiley and Sons, 2014 (ISBN 047150338X, 9780471503385), Current Protocols in Protein Science (CPPS), John E. Coligan (ed.), John Wiley and Sons, Inc., 2005; and Current Protocols in Immunology (CPI) (John E. Coligan, ADA M Kruisbeek, David H Margulies, Ethan M Shevach, Warren Strobe, (eds.) John Wiley and Sons, Inc., 2003 (ISBN 0471142735, 9780471142737), the contents of which are all incorporated by reference herein in their entireties.

[0076] As used herein, the terms "heterologous nucleotide sequence" and "transgene" are used interchangeably and refer to a nucleic acid of interest (other than a nucleic acid encoding a capsid polypeptide) that is incorporated into and may be delivered and expressed by a ceDNA vector as disclosed herein. Transgenes of interest include, but are not limited to, nucleic acids encoding polypeptides, preferably therapeutic (*e.g.*, for medical, diagnostic, or veterinary uses) or immunogenic polypeptides (*e.g.*, for vaccines). In some embodiments, nucleic acids of interest include nucleic acids that are transcribed into therapeutic RNA. Transgenes included for use in the ceDNA vectors of the invention include, but are not limited to, those that express or encode one or more polypeptides, peptides, ribozymes, aptamers, peptide nucleic acids, siRNAs, RNAis, miRNAs, lncRNAs, antisense oligo- or polynucleotides, antibodies, antigen binding fragments, or any combination thereof.

[0077] As used herein, the terms "expression cassette" and "transcription cassette" are used interchangeably and refer to a linear stretch of nucleic acids that includes a transgene that is operably linked to one or more promoters or other regulatory sequences sufficient to direct transcription of the transgene, but which does not comprise capsid-encoding sequences, other vector sequences or inverted terminal repeat

regions. An expression cassette may additionally comprise one or more *cis*-acting sequences (*e.g.*, promoters, enhancers, or repressors), one or more introns, and one or more post-transcriptional regulatory elements.

[0078] As used herein, the term "terminal repeat" or "TR" includes any viral terminal repeat or synthetic sequence that comprises at least one minimal required origin of replication and a region comprising a palindrome hairpin structure. A Rep-binding sequence ("RBS") (also referred to as RBE (Rep-binding element)) and a terminal resolution site ("TRS") together constitute a "minimal required origin of replication" and thus the TR comprises at least one RBS and at least one TRS. TRs that are the inverse complement of one another within a given stretch of polynucleotide sequence are typically each referred to as an "inverted terminal repeat" or "ITR". In the context of a virus, ITRs mediate replication, virus packaging, integration and provirus rescue. As was unexpectedly found in the invention herein, TRs that are not inverse complements across their full length can still perform the traditional functions of ITRs, and thus the term ITR is used herein to refer to a TR in a ceDNA genome or ceDNA vector that is capable of mediating replication of ceDNA vector. It will be understood by one of ordinary skill in the art that in complex ceDNA vector configurations more than two ITRs or asymmetric ITR pairs may be present. The ITR can be an AAV ITR or a non-AAV ITR, or can be derived from an AAV ITR or a non-AAV ITR. For example, the ITR can be derived from the family Parvoviridae, which encompasses parvoviruses and dependoviruses (*e.g.*, canine parvovirus, bovine parvovirus, mouse parvovirus, porcine parvovirus, human parvovirus B-19), or the SV40 hairpin that serves as the origin of SV40 replication can be used as an ITR, which can further be modified by truncation, substitution, deletion, insertion and/or addition. Parvoviridae family viruses consist of two subfamilies: Parvovirinae, which infect vertebrates, and Densovirinae, which infect invertebrates. Dependoparvoviruses include the viral family of the adeno-associated viruses (AAV) which are capable of replication in vertebrate hosts including, but not limited to, human, primate, bovine, canine, equine and ovine species. For convenience herein, an ITR located 5' to (upstream of) an expression cassette in a ceDNA vector is referred to as a "5' ITR" or a "left ITR", and an ITR located 3' to (downstream of) an expression cassette in a ceDNA vector is referred to as a "3' ITR" or a "right ITR".

[0079] A "wild-type ITR" or "WT-ITR" refers to the sequence of a naturally occurring ITR sequence in an AAV or other dependovirus that retains, *e.g.*, Rep binding activity and Rep nicking ability. The nucleotide sequence of a WT-ITR from any AAV serotype may slightly vary from the canonical naturally occurring sequence due to degeneracy of the genetic code or drift, and therefore WT-ITR sequences encompassed for use herein include WT-ITR sequences as result of naturally occurring changes taking place during the production process (*e.g.*, a replication error).

[0080] As used herein, the term "substantially symmetrical WT-ITRs" or a "substantially symmetrical WT-ITR pair" refers to a pair of WT-ITRs within a single ceDNA genome or ceDNA vector that are both wild type ITRs that have an inverse complement sequence across their entire length. For example, an ITR can be considered to be a wild-type sequence, even if it has one or more nucleotides that deviate from the canonical

naturally occurring sequence, so long as the changes do not affect the properties and overall three-dimensional structure of the sequence. In some aspects, the deviating nucleotides represent conservative sequence changes. As one non-limiting example, a sequence that has at least 95%, 96%, 97%, 98%, or 99% sequence identity to the canonical sequence (as measured, *e.g.*, using BLAST at default settings), and also has a symmetrical three-dimensional spatial organization to the other WT-ITR such that their 3D structures are the same shape in geometrical space. The substantially symmetrical WT-ITR has the same A, C-C' and B-B' loops in 3D space. A substantially symmetrical WT-ITR can be functionally confirmed as WT by determining that it has an operable Rep binding site (RBE or RBE') and terminal resolution site (trs) that pairs with the appropriate Rep protein. One can optionally test other functions, including transgene expression under permissive conditions.

[0081] As used herein, the phrases of "modified ITR" or "mod-ITR" or "mutant ITR" are used interchangeably herein and refer to an ITR that has a mutation in at least one or more nucleotides as compared to the WT-ITR from the same serotype. The mutation can result in a change in one or more of A, C, C', B, B' regions in the ITR, and can result in a change in the three-dimensional spatial organization (*i.e.* its 3D structure in geometric space) as compared to the 3D spatial organization of a WT-ITR of the same serotype.

[0082] As used herein, the term "asymmetric ITRs" also referred to as "asymmetric ITR pairs" refers to a pair of ITRs within a single ceDNA genome or ceDNA vector that are not inverse complements across their full length. As one non-limiting example, an asymmetric ITR pair does not have a symmetrical three-dimensional spatial organization to their cognate ITR such that their 3D structures are different shapes in geometrical space. Stated differently, an asymmetrical ITR pair have the different overall geometric structure, *i.e.*, they have different organization of their A, C-C' and B-B' loops in 3D space (*e.g.*, one ITR may have a short C-C' arm and/or short B-B' arm as compared to the cognate ITR). The difference in sequence between the two ITRs may be due to one or more nucleotide addition, deletion, truncation, or point mutation. In one embodiment, one ITR of the asymmetric ITR pair may be a wild-type AAV ITR sequence and the other ITR a modified ITR as defined herein (*e.g.*, a non-wild-type or synthetic ITR sequence). In another embodiment, neither ITRs of the asymmetric ITR pair is a wild-type AAV sequence and the two ITRs are modified ITRs that have different shapes in geometrical space (*i.e.*, a different overall geometric structure). In some embodiments, one mod-ITRs of an asymmetric ITR pair can have a short C-C' arm and the other ITR can have a different modification (*e.g.*, a single arm, or a short B-B' arm etc.) such that they have different three-dimensional spatial organization as compared to the cognate asymmetric mod-ITR.

[0083] As used herein, the term "symmetric ITRs" refers to a pair of ITRs within a single ceDNA genome or ceDNA vector that are mutated or modified relative to wild-type dependoviral ITR sequences and are inverse complements across their full length. Neither ITRs are wild type ITR AAV2 sequences (*i.e.*, they are a modified ITR, also referred to as a mutant ITR), and can have a difference in sequence from the wild

type ITR due to nucleotide addition, deletion, substitution, truncation, or point mutation. For convenience herein, an ITR located 5' to (upstream of) an expression cassette in a ceDNA vector is referred to as a "5' ITR" or a "left ITR", and an ITR located 3' to (downstream of) an expression cassette in a ceDNA vector is referred to as a "3' ITR" or a "right ITR".

[0084] As used herein, the terms "substantially symmetrical modified-ITRs" or a "substantially symmetrical mod-ITR pair" refers to a pair of modified-ITRs within a single ceDNA genome or ceDNA vector that are both that have an inverse complement sequence across their entire length. For example, the a modified ITR can be considered substantially symmetrical, even if it has some nucleotide sequences that deviate from the inverse complement sequence so long as the changes do not affect the properties and overall shape. As one non-limiting example, a sequence that has at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to the canonical sequence (as measured using BLAST at default settings), and also has a symmetrical three-dimensional spatial organization to their cognate modified ITR such that their 3D structures are the same shape in geometrical space. Stated differently, a substantially symmetrical modified-ITR pair have the same A, C-C' and B-B' loops organized in 3D space. In some embodiments, the ITRs from a mod-ITR pair may have different reverse complement nucleotide sequences but still have the same symmetrical three-dimensional spatial organization – that is both ITRs have mutations that result in the same overall 3D shape. For example, one ITR (*e.g.*, 5' ITR) in a mod-ITR pair can be from one serotype, and the other ITR (*e.g.*, 3' ITR) can be from a different serotype, however, both can have the same corresponding mutation (*e.g.*, if the 5'ITR has a deletion in the C region, the cognate modified 3'ITR from a different serotype has a deletion at the corresponding position in the C' region), such that the modified ITR pair has the same symmetrical three-dimensional spatial organization. In such embodiments, each ITR in a modified ITR pair can be from different serotypes (*e.g.* AAV1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12) such as the combination of AAV2 and AAV6, with the modification in one ITR reflected in the corresponding position in the cognate ITR from a different serotype. In one embodiment, a substantially symmetrical modified ITR pair refers to a pair of modified ITRs (mod-ITRs) so long as the difference in nucleotide sequences between the ITRs does not affect the properties or overall shape and they have substantially the same shape in 3D space. As a non-limiting example, a mod-ITR that has at least 95%, 96%, 97%, 98% or 99% sequence identity to the canonical mod-ITR as determined by standard means well known in the art such as BLAST (Basic Local Alignment Search Tool), or BLASTN at default settings, and also has a symmetrical three-dimensional spatial organization such that their 3D structure is the same shape in geometric space. A substantially symmetrical mod-ITR pair has the same A, C-C' and B-B' loops in 3D space, *e.g.*, if a modified ITR in a substantially symmetrical mod-ITR pair has a deletion of a C-C' arm, then the cognate mod-ITR has the corresponding deletion of the C-C' loop and also has a similar 3D structure of the remaining A and B-B' loops in the same shape in geometric space of its cognate mod-ITR.

[0085] The term "flanking" refers to a relative position of one nucleic acid sequence with respect to another nucleic acid sequence. Generally, in the sequence ABC, B is flanked by A and C. The same is true for the arrangement AxBxC. Thus, a flanking sequence precedes or follows a flanked sequence but need not be

contiguous with, or immediately adjacent to the flanked sequence. In one embodiment, the term flanking refers to terminal repeats at each end of the linear duplex ceDNA vector.

[0086] As used herein, the term "ceDNA genome" refers to an expression cassette that further incorporates at least one inverted terminal repeat region. A ceDNA genome may further comprise one or more spacer regions. In some embodiments the ceDNA genome is incorporated as an intermolecular duplex polynucleotide of DNA into a plasmid or viral genome.

[0087] As used herein, the term "ceDNA spacer region" refers to an intervening sequence that separates functional elements in the ceDNA vector or ceDNA genome. In some embodiments, ceDNA spacer regions keep two functional elements at a desired distance for optimal functionality. In some embodiments, ceDNA spacer regions provide or add to the genetic stability of the ceDNA genome within *e.g.*, a plasmid or baculovirus. In some embodiments, ceDNA spacer regions facilitate ready genetic manipulation of the ceDNA genome by providing a convenient location for cloning sites and the like. For example, in certain aspects, an oligonucleotide "polylinker" containing several restriction endonuclease sites, or a non-open reading frame sequence designed to have no known protein (*e.g.*, transcription factor) binding sites can be positioned in the ceDNA genome to separate the *cis*-acting factors, *e.g.*, inserting a 6mer, 12mer, 18mer, 24mer, 48mer, 86mer, 176mer, *etc*. between the terminal resolution site and the upstream transcriptional regulatory element. Similarly, the spacer may be incorporated between the polyadenylation signal sequence and the 3'-terminal resolution site.

[0088] As used herein, the terms "Rep binding site, "Rep binding element, "RBE" and "RBS" are used interchangeably and refer to a binding site for Rep protein (*e.g.*, AAV Rep 78 or AAV Rep 68) which upon binding by a Rep protein permits the Rep protein to perform its site-specific endonuclease activity on the sequence incorporating the RBS. An RBS sequence and its inverse complement together form a single RBS. RBS sequences are known in the art, and include, for example, 5'-GCGCGCTCGCTCGCTC-3' (SEQ ID NO: 531), an RBS sequence identified in AAV2. Any known RBS sequence may be used in the embodiments of the invention, including other known AAV RBS sequences and other naturally known or synthetic RBS sequences. Without being bound by theory it is thought that he nuclease domain of a Rep protein binds to the duplex nucleotide sequence GCTC, and thus the two known AAV Rep proteins bind directly to and stably assemble on the duplex oligonucleotide, 5'-(GCGC)(GCTC)(GCTC)(GCTC)-3' (SEQ ID NO: 531). In addition, soluble aggregated conformers (*i.e.*, undefined number of inter-associated Rep proteins) dissociate and bind to oligonucleotides that contain Rep binding sites. Each Rep protein interacts with both the nitrogenous bases and phosphodiester backbone on each strand. The interactions with the nitrogenous bases provide sequence specificity whereas the interactions with the phosphodiester backbone are non- or less- sequence specific and stabilize the protein-DNA complex.

[0089] As used herein, the terms "terminal resolution site" and "TRS" are used interchangeably herein and refer to a region at which Rep forms a tyrosine-phosphodiester bond with the 5' thymidine generating a 3' OH that serves as a substrate for DNA extension via a cellular DNA polymerase, *e.g.*, DNA pol delta or

DNA pol epsilon. Alternatively, the Rep-thymidine complex may participate in a coordinated ligation reaction. In some embodiments, a TRS minimally encompasses a non-base-paired thymidine. In some embodiments, the nicking efficiency of the TRS can be controlled at least in part by its distance within the same molecule from the RBS. When the acceptor substrate is the complementary ITR, then the resulting product is an intramolecular duplex. TRS sequences are known in the art, and include, for example, 5'-GGTTGA-3' (SEQ ID NO: 45), the hexanucleotide sequence identified in AAV2. Any known TRS sequence may be used in the embodiments of the invention, including other known AAV TRS sequences and other naturally known or synthetic TRS sequences such as AGTT (SEQ ID NO: 46), GGTTGG (SEQ ID NO: 47), AGTTGG (SEQ ID NO: 48), AGTTGA (SEQ ID NO: 49), and other motifs such as RRTTRR (SEQ ID NO: 50).

[0090] As used herein, the term "ceDNA-plasmid" refers to a plasmid that comprises a ceDNA genome as an intermolecular duplex.

[0091] As used herein, the term "ceDNA-bacmid" refers to an infectious baculovirus genome comprising a ceDNA genome as an intermolecular duplex that is capable of propagating in *E. coli* as a plasmid, and so can operate as a shuttle vector for baculovirus.

[0092] As used herein, the term "ceDNA-baculovirus" refers to a baculovirus that comprises a ceDNA genome as an intermolecular duplex within the baculovirus genome.

[0093] As used herein, the terms "ceDNA-baculovirus infected insect cell" and "ceDNA-BIIC" are used interchangeably, and refer to an invertebrate host cell (including, but not limited to an insect cell (*e.g.*, an Sf9 cell)) infected with a ceDNA-baculovirus.

[0094] As used herein, the term "ceDNA" refers to capsid-free closed-ended linear double stranded (ds) duplex DNA for non-viral gene transfer, synthetic or otherwise. Detailed description of ceDNA is described in International application of PCT/US2017/020828, filed March 3, 2017, the entire contents of which are expressly incorporated herein by reference. Certain methods for the production of ceDNA comprising various inverted terminal repeat (ITR) sequences and configurations using cell-based methods are described in Example 1 of International applications PCT/US18/49996, filed September 7, 2018, and PCT/US2018/064242, filed December 6, 2018 each of which is incorporated herein in its entirety by reference. Certain methods for the production of synthetic ceDNA vectors comprising various ITR sequences and configurations are described, *e.g.*, in International application PCT/US2019/14122, filed January 18, 2019, the entire content of which is incorporated herein by reference.

[0095] As used herein, the term "closed-ended DNA vector" refers to a capsid-free DNA vector with at least one covalently closed end and where at least part of the vector has an intramolecular duplex structure.

[0096] As used herein, the terms "ceDNA vector" and "ceDNA" are used interchangeably and refer to a closed-ended DNA vector comprising at least one terminal palindrome. In some embodiments, the ceDNA comprises two covalently-closed ends.

[0097] As used herein, the term "neDNA" or "nicked ceDNA" refers to a closed-ended DNA having a nick or a gap of 1-100 base pairs in a stem region or spacer region 5' upstream of an open reading frame (*e.g.*, a promoter and transgene to be expressed).

[0098] As used herein, the terms "gap" and "nick" are used interchangeably and refer to a discontinued portion of synthetic DNA vector of the present invention, creating a stretch of single stranded DNA portion in otherwise double stranded ceDNA. The gap can be 1 base-pair to 100 base-pair long in length in one strand of a duplex DNA. Typical gaps, designed and created by the methods described herein and synthetic vectors generated by the methods can be, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59 or 60 bp long in length. Exemplified gaps in the present disclosure can be 1 bp to 10 bp long, 1 to 20 bp long, 1 to 30 bp long in length.

[0099] As defined herein, "reporters" refer to proteins that can be used to provide detectable read-outs. Reporters generally produce a measurable signal such as fluorescence, color, or luminescence. Reporter protein coding sequences encode proteins whose presence in the cell or organism is readily observed. For example, fluorescent proteins cause a cell to fluoresce when excited with light of a particular wavelength, luciferases cause a cell to catalyze a reaction that produces light, and enzymes such as β-galactosidase convert a substrate to a colored product. Exemplary reporter polypeptides useful for experimental or diagnostic purposes include, but are not limited to β-lactamase, β -galactosidase (LacZ), alkaline phosphatase (AP), thymidine kinase (TK), green fluorescent protein (GFP) and other fluorescent proteins, chloramphenicol acetyltransferase (CAT), luciferase, and others well known in the art.

[00100] As used herein, the term "effector protein" refers to a polypeptide that provides a detectable read-out, either as, for example, a reporter polypeptide, or more appropriately, as a polypeptide that kills a cell, *e.g.*, a toxin, or an agent that renders a cell susceptible to killing with a chosen agent or lack thereof. Effector proteins include any protein or peptide that directly targets or damages the host cell's DNA and/or RNA. For example, effector proteins can include, but are not limited to, a restriction endonuclease that targets a host cell DNA sequence (whether genomic or on an extrachromosomal element), a protease that degrades a polypeptide target necessary for cell survival, a DNA gyrase inhibitor, and a ribonuclease-type toxin. In some embodiments, the expression of an effector protein controlled by a synthetic biological circuit as described herein can participate as a factor in another synthetic biological circuit to thereby expand the range and complexity of a biological circuit system's responsiveness.

[00101] Transcriptional regulators refer to transcriptional activators and repressors that either activate or repress transcription of a gene of interest. Promoters are regions of nucleic acid that initiate transcription of a particular gene Transcriptional activators typically bind nearby to transcriptional promoters and recruit RNA polymerase to directly initiate transcription. Repressors bind to transcriptional promoters and sterically hinder transcriptional initiation by RNA polymerase. Other transcriptional regulators may serve as either an activator or a repressor depending on where they bind and cellular and environmental conditions.

Non-limiting examples of transcriptional regulator classes include, but are not limited to homeodomain proteins, zinc-finger proteins, winged-helix (forkhead) proteins, and leucine-zipper proteins.

[00102] As used herein, a "repressor protein" or "inducer protein" is a protein that binds to a regulatory sequence element and represses or activates, respectively, the transcription of sequences operatively linked to the regulatory sequence element. Preferred repressor and inducer proteins as described herein are sensitive to the presence or absence of at least one input agent or environmental input. Preferred proteins as described herein are modular in form, comprising, for example, separable DNA-binding and input agent-binding or responsive elements or domains.

[00103] As used herein, "carrier" includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutically active substances is well known in the art. Supplementary active ingredients can also be incorporated into the compositions. The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce a toxic, an allergic, or similar untoward reaction when administered to a host.

[00104] As used herein, an "input agent responsive domain" is a domain of a transcription factor that binds to or otherwise responds to a condition or input agent in a manner that renders a linked DNA binding fusion domain responsive to the presence of that condition or input. In one embodiment, the presence of the condition or input results in a conformational change in the input agent responsive domain, or in a protein to which it is fused, that modifies the transcription-modulating activity of the transcription factor.

[00105] The term "*in vivo*" refers to assays or processes that occur in or within an organism, such as a multicellular animal. In some of the aspects described herein, a method or use can be said to occur "*in vivo*" when a unicellular organism, such as a bacterium, is used. The term "*ex vivo*" refers to methods and uses that are performed using a living cell with an intact membrane that is outside of the body of a multicellular animal or plant, *e.g.*, explants, cultured cells, including primary cells and cell lines, transformed cell lines, and extracted tissue or cells, including blood cells, among others. The term "*in vitro*" refers to assays and methods that do not require the presence of a cell with an intact membrane, such as cellular extracts, and can refer to the introducing of a programmable synthetic biological circuit in a non-cellular system, such as a medium not comprising cells or cellular systems, such as cellular extracts.

[00106] The term "promoter," as used herein, refers to any nucleic acid sequence that regulates the expression of another nucleic acid sequence by driving transcription of the nucleic acid sequence, which can be a heterologous target gene encoding a protein or an RNA. Promoters can be constitutive, inducible, repressible, tissue-specific, or any combination thereof. A promoter is a control region of a nucleic acid sequence at which initiation and rate of transcription of the remainder of a nucleic acid sequence are controlled. A promoter can also contain genetic elements at which regulatory proteins and molecules can bind, such as RNA polymerase and other transcription factors. In some embodiments of the aspects described herein, a promoter can drive the expression of a transcription factor that regulates the expression

of the promoter itself, or that of another promoter used in another modular component of the synthetic biological circuits described herein. Within the promoter sequence will be found a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Various promoters, including inducible promoters, may be used to drive the expression of transgenes in the ceDNA vectors disclosed herein.

[00107] The term "enhancer" as used herein refers a cis-acting regulatory sequence (*e.g.*, 50-1,500 base pairs) that bind one or more proteins (*e.g.*, activator proteins, or transcription factor) to increase transcriptional activation of a nucleic acid sequence. Enhancers can be positioned up to 1,000,000 base pars upstream of the gene start site or downstream of the gene start site that they regulate. An enhancer can be positioned within an intronic region, or in the exonic region of an unrelated gene.

[00108] A promoter can be said to drive expression or drive transcription of the nucleic acid sequence that it regulates. The phrases "operably linked," "operatively positioned," "operatively linked," "under control," and "under transcriptional control" indicate that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence it regulates to control transcriptional initiation and/or expression of that sequence. An "inverted promoter," as used herein, refers to a promoter in which the nucleic acid sequence is in the reverse orientation, such that what was the coding strand is now the non-coding strand, and vice versa. Inverted promoter sequences can be used in various embodiments to regulate the state of a switch. In addition, in various embodiments, a promoter can be used in conjunction with an enhancer.

[00109] A promoter can be one naturally associated with a gene or sequence, as can be obtained by isolating the 5' non-coding sequences located upstream of the coding segment and/or exon of a given gene or sequence. Such a promoter can be referred to as "endogenous." Similarly, in some embodiments, an enhancer can be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence.

[00110] In some embodiments, a coding nucleic acid segment is positioned under the control of a "recombinant promoter" or "heterologous promoter," both of which refer to a promoter that is not normally associated with the encoded nucleic acid sequence it is operably linked to in its natural environment. A recombinant or heterologous enhancer refers to an enhancer not normally associated with a given nucleic acid sequence in its natural environment. Such promoters or enhancers can include promoters or enhancers of other genes; promoters or enhancers isolated from any other prokaryotic, viral, or eukaryotic cell; and synthetic promoters or enhancers that are not "naturally occurring," *i.e.*, comprise different elements of different transcriptional regulatory regions, and/or mutations that alter expression through methods of genetic engineering that are known in the art. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, promoter sequences can be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR, in connection with the synthetic biological circuits and

modules disclosed herein (see, *e.g.*, U.S. Pat. No. 4,683,202, U.S. Pat. No. 5,928,906, each incorporated herein by reference). Furthermore, it is contemplated that control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts, and the like, can be employed as well.

[00111] As described herein, an "inducible promoter" is one that is characterized by initiating or enhancing transcriptional activity when in the presence of, influenced by, or contacted by an inducer or inducing agent. An "inducer" or "inducing agent," as defined herein, can be endogenous, or a normally exogenous compound or protein that is administered in such a way as to be active in inducing transcriptional activity from the inducible promoter. In some embodiments, the inducer or inducing agent, *i.e.*, a chemical, a compound or a protein, can itself be the result of transcription or expression of a nucleic acid sequence (*i.e.*, an inducer can be an inducer protein expressed by another component or module), which itself can be under the control or an inducible promoter. In some embodiments, an inducible promoter is induced in the absence of certain agents, such as a repressor. Examples of inducible promoters include but are not limited to, tetracycline, metallothionine, ecdysone, mammalian viruses (*e.g.*, the adenovirus late promoter; and the mouse mammary tumor virus long terminal repeat (MMTV-LTR)) and other steroid-responsive promoters, rapamycin responsive promoters and the like.

[00112] The term "subject" as used herein refers to a human or animal, to whom treatment, including prophylactic treatment, with the ceDNA vector according to the present invention, is provided. Usually the animal is a vertebrate such as, but not limited to a primate, rodent, domestic animal or game animal. Primates include but are not limited to, chimpanzees, cynomologous monkeys, spider monkeys, and macaques, *e.g.*, Rhesus. Rodents include mice, rats, woodchucks, ferrets, rabbits and hamsters. Domestic and game animals include, but are not limited to, cows, horses, pigs, deer, bison, buffalo, feline species, *e.g.*, domestic cat, canine species, *e.g.*, dog, fox, wolf, avian species, *e.g.*, chicken, emu, ostrich, and fish, *e.g.*, trout, catfish and salmon. In certain embodiments of the aspects described herein, the subject is a mammal, *e.g.*, a primate or a human. A subject can be male or female. Additionally, a subject can be an infant or a child. In some embodiments, the subject can be a neonate or an unborn subject, *e.g.*, the subject is *in utero.* Preferably, the subject is a mammal. The mammal can be a human, non-human primate, mouse, rat, dog, cat, horse, or cow, but is not limited to these examples. Mammals other than humans can be advantageously used as subjects that represent animal models of diseases and disorders. In addition, the methods and compositions described herein can be used for domesticated animals and/or pets. A human subject can be of any age, gender, race or ethnic group, *e.g.*, Caucasian (white), Asian, African, black, African American, African European, Hispanic, Mideastern, etc. In some embodiments, the subject can be a patient or other subject in a clinical setting. In some embodiments, the subject is already undergoing treatment.

[00113] As used herein, the term "antibody" is used in the broadest sense and encompasses various antibody structures, including but not limited to monoclonal antibodies, polyclonal antibodies, multispecific antibodies (*e.g.*, bispecific antibodies), and antibody fragments so long as they exhibit the desired antigen-

binding activity. An "antibody fragment" refers to a molecule other than an intact antibody that comprises a portion of an intact antibody that binds the same antigen to which the intact antibody binds. In one embodiment, the antibody or antibody fragment comprises an immunoglobulin chain or antibody fragment and at least one immunoglobulin variable domain sequence. Examples of antibodies or fragments thereof include, but are not limited to, an Fv, an scFv, a Fab fragment, a Fab', a $F(ab')_2$, a Fab'-SH, a single domain antibody (dAb), a heavy chain, a light chain, a heavy and light chain, a full antibody (*e.g.*, includes each of the Fc, Fab, heavy chains, light chains, variable regions etc.), a bispecific antibody, a diabody, a linear antibody, a single chain antibody, an intrabody, a monoclonal antibody, a chimeric antibody, a multispecific antibody, or a multimeric antibody. An antibody or fragment thereof can be of any class, including but not limited to IgA, IgD, IgE, IgG, and IgM, and of any subclass thereof including but not limited to IgG1, IgG2, IgG3, IgG4, IgA1 and IgA2. In addition, an antibody can be derived from any mammal, for example, primates, humans, rats, mice, horses, goats etc. In one embodiment, the antibody is human or humanized. In some embodiments, the antibody is a modified antibody. In some embodiments, the components of an antibody can be expressed separately such that the antibody self-assembles following expression of the protein components. In some embodiments, the antibody is "humanized" to reduce immunogenic reactions in a human. In some embodiments, the antibody has a desired function, for example, interaction and inhibition of a desired protein for the purpose of treating a disease or a symptom of a disease. In one embodiment, the antibody or antibody fragment comprises a framework region or an F_c region.

[00114] As used herein, the term "antigen-binding domain" of an antibody molecule refers to the part of an antibody molecule, *e.g.*, an immunoglobulin (Ig) molecule, that participates in antigen binding. In embodiments, the antigen binding site is formed by amino acid residues of the variable (V) regions of the heavy (H) and light (L) chains. Three highly divergent stretches within the variable regions of the heavy and light chains, referred to as hypervariable regions, are disposed between more conserved flanking stretches called "framework regions," (FRs). FRs are amino acid sequences that are naturally found between, and adjacent to, hypervariable regions in immunoglobulins. In embodiments, in an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface, which is complementary to the three-dimensional surface of a bound antigen. The three hypervariable regions of each of the heavy and light chains are referred to as "complementarity-determining regions," or "CDRs." The framework region and CDRs have been defined and described, *e.g.*, in Kabat, E. A., *et al.* (1991) Sequences of Proteins of Immunological Interest, Fifth Edition, U.S. Department of Health and Human Services, NIH Publication No. 91-3242, and Chothia, C. *et al.* (1987) J. Mol. Biol. 196:901-917. Each variable chain (*e.g.*, variable heavy chain and variable light chain) is typically made up of three CDRs and four FRs, arranged from amino-terminus to carboxy-terminus in the amino acid order: FR1, CDR1, FR2, CDR2, FR3, CDR3, and FR4.

[00115] As used herein, the term "full length antibody" refers to an immunoglobulin (Ig) molecule (*e.g.*, an IgG antibody), for example, that is naturally occurring, and formed by normal immunoglobulin gene fragment recombinatorial processes.

[00116] As used herein, the term "functional antibody fragment" refers to a fragment that binds to the same antigen as that recognized by the intact (*e.g.*, full-length) antibody. The terms "antibody fragment" or "functional fragment" also include isolated fragments consisting of the variable regions, such as the "Fv" fragments consisting of the variable regions of the heavy and light chains or recombinant single chain polypeptide molecules in which light and heavy variable regions are connected by a peptide linker ("scFv proteins"). In some embodiments, an antibody fragment does not include portions of antibodies without antigen binding activity, such as Fc fragments or single amino acid residues.

[00117] As used herein, an "immunoglobulin variable domain sequence" refers to an amino acid sequence which can form the structure of an immunoglobulin variable domain. For example, the sequence may include all or part of the amino acid sequence of a naturally-occurring variable domain. For example, the sequence may or may not include one, two, or more N- or C-terminal amino acids, or may include other alterations that are compatible with formation of the protein structure.

[00118] The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Thus, this term includes single, double, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer including purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. "Oligonucleotide" generally refers to polynucleotides of between about 5 and about 100 nucleotides of single- or double-stranded DNA. However, for the purposes of this disclosure, there is no upper limit to the length of an oligonucleotide. Oligonucleotides are also known as "oligomers" or "oligos" and may be isolated from genes, or chemically synthesized by methods known in the art. The terms "polynucleotide" and "nucleic acid" should be understood to include, as applicable to the embodiments being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides. DNA may be in the form of, *e.g.*, antisense molecules, plasmid DNA, DNA-DNA duplexes, pre-condensed DNA, PCR products, vectors (P1, PAC, BAC, YAC, artificial chromosomes), expression cassettes, chimeric sequences, chromosomal DNA, or derivatives and combinations of these groups. DNA may be in the form of minicircle, plasmid, bacmid, minigene, ministring DNA (linear covalently closed DNA vector), closed-ended linear duplex DNA (CELiD or ceDNA), doggybone (dbDNA ™) DNA, dumbbell shaped DNA, minimalistic immunological-defined gene expression (MIDGE)-vector, viral vector or nonviral vectors. RNA may be in the form of small interfering RNA (siRNA), Dicer-substrate dsRNA, small hairpin RNA (shRNA), asymmetrical interfering RNA (aiRNA), microRNA (miRNA), mRNA, rRNA, tRNA, viral RNA (vRNA), and combinations thereof. Nucleic acids include nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, and which have similar binding properties as the

reference nucleic acid. Examples of such analogs and/or modified residues include, without limitation, phosphorothioates, phosphorodiamidate morpholino oligomer (morpholino), phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2'-O-methyl ribonucleotides, locked nucleic acid (LNA™), and peptide nucleic acids (PNAs). Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions), alleles, orthologs, SNPs, and complementary sequences as well as the sequence explicitly indicated.

[00119] "Nucleotides" contain a sugar deoxyribose (DNA) or ribose (RNA), a base, and a phosphate group. Nucleotides are linked together through the phosphate groups.

[00120] "Bases" include purines and pyrimidines, which further include natural compounds adenine, thymine, guanine, cytosine, uracil, inosine, and natural analogs, and synthetic derivatives of purines and pyrimidines, which include, but are not limited to, modifications which place new reactive groups such as, but not limited to, amines, alcohols, thiols, carboxylates, and alkylhalides.

[00121] By "hybridizable" or "complementary" or "substantially complementary" it is meant that a nucleic acid (*e.g.,* RNA) includes a sequence of nucleotides that enables it to non-covalently bind, *i.e.* form Watson-Crick base pairs and/or G/U base pairs, "anneal", or "hybridize," to another nucleic acid in a sequence-specific, antiparallel, manner (*i.e.,* a nucleic acid specifically binds to a complementary nucleic acid) under the appropriate *in vitro* and/or *in vivo* conditions of temperature and solution ionic strength. As is known in the art, standard Watson-Crick base-pairing includes: adenine (A) pairing with thymidine (T), adenine (A) pairing with uracil (U), and guanine (G) pairing with cytosine (C). In addition, it is also known in the art that for hybridization between two RNA molecules (*e.g.,* dsRNA), guanine (G) base pairs with uracil (U). For example, G/U base-pairing is partially responsible for the degeneracy (*i.e.,* redundancy) of the genetic code in the context of tRNA anti-codon base-pairing with codons in mRNA. In the context of this disclosure, a guanine (G) of a protein-binding segment (dsRNA duplex) of a subject DNA-targeting RNA molecule is considered complementary to a uracil (U), and vice versa. As such, when a G/U base-pair can be made at a given nucleotide position a protein-binding segment (dsRNA duplex) of a subject DNA-targeting RNA molecule, the position is not considered to be non-complementary, but is instead considered to be complementary.

[00122] The term "nucleic acid construct" as used herein refers to a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which is modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature or which is synthetic. The term nucleic acid construct is synonymous with the term "expression cassette" when the nucleic acid construct contains the control sequences required for expression of a coding sequence of the present disclosure. An "expression cassette" includes a DNA coding sequence operably linked to a promoter.

[00123] As used herein, the phrases "nucleic acid therapeutic", "therapeutic nucleic acid" and "TNA" are used interchangeably and refer to any modality of therapeutic using nucleic acids as an active component of therapeutic agent to treat a disease or disorder. As used herein, these phrases refer to RNA-based therapeutics and DNA-based therapeutics. Non-limiting examples of RNA-based therapeutics include mRNA, antisense RNA and oligonucleotides, ribozymes, aptamers, interfering RNAs (RNAi), Dicer-substrate dsRNA, small hairpin RNA (shRNA), asymmetrical interfering RNA (aiRNA), microRNA (miRNA). Non-limiting examples of DNA-based therapeutics include minicircle DNA, minigene, viral DNA (*e.g.*, Lentiviral or AAV genome) or non-viral synthetic DNA vectors, closed-ended linear duplex DNA (ceDNA / CELiD), plasmids, bacmids, doggybone (dbDNA™) DNA vectors, minimalistic immunological-defined gene expression (MIDGE)-vector, nonviral ministring DNA vector (linear-covalently closed DNA vector), or dumbbell-shaped DNA minimal vector ("dumbbell DNA").

[00124] The terms "peptide," "polypeptide," and "protein" are used interchangeably herein, and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.

[00125] As used herein, the term "synthetic AAV vector" and "synthetic production of AAV vector" refers to an AAV vector and synthetic production methods thereof in an entirely cell-free environment.

[00126] As used herein the term "comprising" or "comprises" is used in reference to compositions, methods, and respective component(s) thereof, that are essential to the method or composition, yet open to the inclusion of unspecified elements, whether essential or not.

[00127] As used herein the term "consisting essentially of" refers to those elements required for a given embodiment. The term permits the presence of elements that do not materially affect the basic and novel or functional characteristic(s) of that embodiment.

[00128] The term "consisting of" refers to compositions, methods, and respective components thereof as described herein, which are exclusive of any element not recited in that description of the embodiment.

[00129] As used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural references unless the context clearly dictates otherwise. Thus, for example, references to "the method" includes one or more methods, and/or steps of the type described herein and/or which will become apparent to those persons skilled in the art upon reading this disclosure and so forth. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of this disclosure, suitable methods and materials are described below. The abbreviation, "*e.g.*" is derived from the Latin exempli gratia, and is used herein to indicate a non-limiting example. Thus, the abbreviation "*e.g.*" is synonymous with the term "for example."

[00130] Other than in the operating examples, or where otherwise indicated, all numbers expressing quantities of ingredients or reaction conditions used herein should be understood as modified in all instances

by the term "about." The term "about" when used in connection with percentages can mean ±1%. The present invention is further explained in detail by the following examples, but the scope of the invention should not be limited thereto.

[00131] It should be understood that this invention is not limited to the particular methodology, protocols, and reagents, etc., described herein and as such can vary. The terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which is defined solely by the claims.

II. Replication initiator (Rep) Proteins

[00132] As described herein, the technology described herein relates to a composition and improved methods of production of DNA vectors, *e.g.*, a ceDNA vector as described herein or an AAV vector with a *single Rep* protein species. According to some aspects, the disclosure provides a method to produce a DNA vector, *e.g.*, a ceDNA vector as described herein, or a an AAV vector using a single Rep protein, wherein the Rep protein is not Rep52 or Rep40. According to some embodiments, the single Rep protein is Rep78. According to some embodiments, the single Rep protein is Rep68. This is an improved and more efficient method of ceDNA vector production which produces superior ceDNA vector yield than the methods described in the prior art which uses two Rep proteins involving Rep78 or 68 and Rep52 or 40(*e.g.*, Rep78 and Rep 52, see **FIG. 32**). Indeed, prior to the instant invention, it was thought that two Rep proteins, one long (e.g., Rep78 or Rep 68) and one short (e.g., Rep52 or Rep40), must be present to produce AAV particles. In particular, it was thought that Rep78 and Rep52 must be present, either as single units or using a single coding sequence for the Rep78 and Rep52 proteins, to produce AAV particles.

[00133] Accordingly, one aspect of the technology described herein relates to a method to produce a DNA vector, *e.g.*, a ceDNA vector as described herein, or a an AAV vector using a single Rep protein, as opposed to two Rep proteins. According to some embodiments, the single Rep protein is Rep78. According to some embodiments, the single Rep protein is Rep68. According to some embodiments, Rep protein can be a Rep78 and Rep68, but not Rep52 or Rep40.

[00134] Another aspect of the technology described herein relates to a composition comprising a nucleic acid construct that comprises a first nucleotide sequence encoding a single parvoviral Rep protein, where the nucleotide sequence does not have an open reading frame (ORF) and lacks a functional initiation codon downstream of the first initiation codon and/or lacks alternative splicing sites preventing exon skipping, thereby enabling the translation of a single parvoviral Rep protein (*e.g.*, a Rep78 or Rep68 protein) without the translation of additional Rep proteins (*e.g.*, any one or more of Rep52 or Rep40) in the insect cells or cell free system. That is, a nucleic acid encoding Rep78 does not also produce a Rep52 protein, and similarly, a nucleic acid encoding Rep68 does not produce a Rep40 protein. Further no other Rep protein is present or expressed in the system. to a nucleic acid construct for the production of DNA vectors, *e.g.*, ceDNA vectors and other recombinant parvovirus (*e.g.* adeno-associated virus) vectors in cells (*e.g.* insect cells, mammalian cells) and cell free systems, where, for example, the insect cells or cell free system.

Rep Proteins in general

[00135] Rep genes function to replicate a viral genome. In wild-type nucleic acid encoding Rep78 or Rep68, a splicing event in the Rep open reading frame of either Rep78 or Rep68 results in two Rep proteins upon translation: Rep52, and Rep40, respectively. That is, Rep78 protein and Rep68 protein are encoded by a single nucleic acid that undergoes differential splicing to produce both Rep 78 and Rep 68. Similarly, Rep 52 protein and Rep 40 protein are encoded by a single nucleic acid that undergoes differential splicing to produce both Rep 52 and Rep 40 proteins. Rep 78 is a full-length protein produced from the original first translation initiation site, whereas Rep52 is a product of translation from a downstream internal "second (AUG)" translation initiation site. Hence, when a full-length wild-type AAV genome is expressed, all four species of Rep proteins are typically present (*e.g.*, Rep78, Rep68, Rep52, and Rep40) largely due to two different translation initiation sites as well as alternative splicing sites present near the carboxy terminus. Rep proteins each comprise various functionalities, for example DNA nicking, DNA binding, helicase, ligase, and ATPase activity. The functionality for a given Rep protein is further described in **FIG. 31**. It has been previously reported that both Rep 78 and Rep 52 proteins are necessary for AAV vector or ceDNA vector production in various systems, *e.g.*, insect cell and mammalian cell systems. However, as discussed herein, the inventors demonstrate that only a single Rep protein, or alternatively at least a combination of long Rep proteins (Rep78 and Rep68), but not short Rep proteins (Rep52 and Rep40), can be used for AAV vector production or ceDNA vector production. The single species of Rep protein useful in the compositions and method as described herein comprises all three functions: DNA nicking, DNA binding and DNA ligation functionality. In certain embodiments, the single Rep protein further comprises helicase and ATPase functionality.

[00136] In some embodiments, the single species of Rep protein useful in the compositions and method as described herein is an AAV2 Rep protein when the ITR is from serotype 2 (*e.g.*, AAV2). In alternative embodiments, a single Rep protein can be from any of the 42 AAV serotypes, or more preferably, from AAV1, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, or AAV12 Rep protein. In some embodiments, a single Rep protein encompassed for use in the methods and compositions as disclosed herein corresponds to an animal parvovirus Rep protein when the ITR is from serotype 2 (*e.g.*, AAV2). The Rep protein works as part of a system with the ITR to bind to the ITR and initiate terminal resolution replication and catalyze the formation of the closed ended ceDNA vector molecule.

[00137] In some embodiments, a single Rep protein useful in the compositions and method as described herein is Rep78. In alternative embodiments, a single Rep protein useful in the compositions and method as described herein or Rep68. In alternative embodiments, a single species of Rep protein is a Rep 52 or Rep40 that has been modified to comprise the functionality of Rep 78 or Rep 68, *e.g.*, to have DNA binding, DNA nicking, helicase, and ATPase activity. Alternatively, in some embodiment, the Rep protein useful in the

composition and method as described herein can be a combination of the long Rep proteins (*e.g.*, Rep78 and Rep68), without Rep52 or Rep40, the short Rep protein(s).

[00138] Another aspect of the technology described herein relates to a nucleic acid construct encoding a single Rep protein, where the nucleic acid does not induce or permit the expression of a second Rep protein. Accordingly, in one aspect, a nucleic acid construct encoding a single Rep protein is modified such that it lacks a functional initiation codon for another Rep protein.

[00139] In one embodiment, the presence of a single Rep species (*e.g.*, with no other species present) is determined by the specific mutations that prevent translation of the p19 Reps, and by absence of other Rep species on western blots using anti-Rep antibodies known in the art.

[00140] *Nucleic acid constructs encoding modified Rep proteins*

[00141] In one embodiment, the single species of Rep protein is encoded by a nucleotide sequence encoding a modified Rep protein, for example, it can encode a modified Rep 78 protein, but the nucleotide sequence does not have a functional initiation codon for encoding the Rep 52 protein, nor does it have the splice sites for exon skipping for production of Rep 68 or Rep40. For example, a modified Rep 78 nucleotide sequence comprises a modification or mutation in the initiation codon for Rep52, such that the initiation codon (*e.g.*, AUG) for Rep52 is changed to no-longer encode methionine, but rather encodes a different amino acid. In some embodiments, the initiation codon (Met) for Rep52 in the Rep78 nucleic acid sequence is mutated to encode glycine (*e.g.*, AUG is mutated to one of: GGU, GGC, GGA, GGG, which encodes Gly), or threonine amino acid (*e.g.* AUG is mutated to one of ACT, ACC, ACA, and ACG, which encodes Thr).

Modified Rep proteins

[00142] In some embodiments, a modified Rep 78 nucleotide sequence can encode a modified Rep 78 protein that comprises a modification of amino acid residue 225 (Met) of SEQ ID NO: 530, wherein the amino acid residue 225 is changed to a glycine (Gly) (e.g, M225G or Met225Gly) or threonine (Thr) (*e.g.*, M225T or Met225Thr). In one embodiment, the mutated Rep 78 protein comprises a sequence of SEQ ID NO: 530, or comprises a sequence having at least 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to SEQ ID NO: 530, where the amino acid at position 225 is not a Met, and where the modified Rep protein has at least DNA binding and DNA nicking functionality, and the gene encoding it does not facilitate production of a second Rep protein. One skilled in the art will be able to generate a point mutation using, *e.g.*, site-directed mutagenesis. To assess if the mutation in the nucleotide sequence was generated correctly, one could perform a sequence alignment with the modified Rep protein (*i.e.*, the Rep protein comprising the point mutation) compared to the wild-type Rep protein.

[00143] In one embodiment, a nucleotide sequence encoding a single Rep protein useful in the compositions and methods as disclosed herein comprises an expression control sequence, *e.g.*, promoter, *cis*-regulatory elements, or regulatory switch as described herein, located upstream of the initiation codon of

the nucleotide sequence encoding the parvoviral Rep78 protein, where the nucleic acid sequence does not have a functional initiation codon for Rep52. In one embodiment, a nucleotide sequence encoding a single Rep protein useful in the compositions and methods as disclosed herein comprises an expression control sequence upstream of the initiation codon of the nucleotide sequence encoding the parvoviral Rep 78 protein, where the nucleic acid sequence does not have a functional spice sites for encoding Rep68.

[00144] That is, in some embodiments, the nucleic acid encoding Rep78 has only one initiation codon, thereby allowing translation of only Rep78 protein or Rep68 protein. In such embodiments, the Rep78 nucleic acid has a functional first initiation codon enabling translation of the Rep78 protein, but the initiation codon downstream of the initial initiation codon is modified (or non-functional) that results in Rep52 not being expressed.

[00145] In all instances no other vectors are used that encode another Rep. Nor is Rep protein already present in the insect cell or mammalian cell used on the methods to generate DNA vectors, *e.g.*, ceDNA vectors or AAV vectors according to the methods as described herein.

[00146] In one embodiment, a single Rep protein useful in the compositions and methods as disclosed herein is from the parvovirus family. In another embodiment, the single Rep protein useful in the compositions and methods as disclosed herein is preferably from a dependovirus subfamily virus Rep. In another embodiment, the single Rep protein useful in the compositions and methods as disclosed herein is more preferably an AAV Rep.

[00147] In one embodiment, a nucleotide sequence of the invention comprises an expression control sequence encoding the AAV Rep 68 protein, where the nucleic acid sequence does not have a functional initiation codon for Rep40, but has a deletion in the intron sequence in its carboxy terminal end, resulting in Rep68. In another embodiment, the nucleic acid sequence has a deletion in the intron sequence of the full-length Rep78 and does not have other functional splice sites resulting in a transcript capable of being translated into Rep 68 only. That is, in some embodiments, the nucleic acid encoding Rep68 has only one initiation codon, thereby allowing translation of only Rep68 protein with the c-terminal intron sequence deleted. In such embodiments, the Rep68 nucleic acid has a functional first initiation codon enabling translation of the Rep68 protein, but the initiation codon downstream of the initial initiation codon is modified or non-functional by a mutation (*e.g.*, M225G or M225T that results in Rep40 not being expressed. Alternatively, a nucleic acid encoding Rep68 is modified such that the second initiation codon is modified or non-functional by a mutation (*e.g.*, M225G or M225T), but the downstream c-terminal splicing sites are operable and allows for expression of the Rep78 protein and Rep68 protein.

[00148] A sequence with substantial identity to the nucleotide sequence of SEQ. ID NO: 530 is a sequence which has at least 60%, 70%, 80% or 90% identity SEQ ID NO: 530.

III. Detailed method of Production of a ceDNA Vector using a single Rep protein

A. Production in General

[00149] As described herein, a ceDNA vector can be obtained by the process using only one Rep protein, as opposed to more than one, *e.g.*, two Rep proteins. Accordingly, one aspect of the present invention relates to a method comprising the steps of: a) incubating a population of host cells (*e.g.* insect cells) harboring the polynucleotide expression construct template (*e.g.*, a ceDNA-plasmid, a ceDNA-Bacmid, and/or a ceDNA-baculovirus), which is devoid of viral capsid coding sequences, in the presence of a single Rep protein under conditions effective and for a time sufficient to induce production of the ceDNA vector within the host cells, and wherein the host cells do not comprise viral capsid coding sequences; and b) harvesting and isolating the ceDNA vector from the host cells. The presence of a single Rep protein induces replication of the vector polynucleotide with a modified ITR to produce the ceDNA vector in a host cell. However, no viral particles (*e.g.* AAV virions) are expressed. Thus, there is no size limitation such as that naturally imposed in AAV or other viral-based vectors.

[00150] The presence of the ceDNA vector isolated from the host cells can be confirmed by digesting DNA isolated from the host cell with a restriction enzyme having a single recognition site on the ceDNA vector and analyzing the digested DNA material on a non-denaturing gel to confirm the presence of characteristic bands of linear and continuous DNA as compared to linear and non-continuous DNA.

[00151] In yet another aspect, the invention provides for use of host cell lines that have stably integrated the DNA vector polynucleotide expression template (ceDNA template) into their own genome in production of the non-viral DNA vector, *e.g.* as described in Lee, L. *et al.* (2013) Plos One 8(8): e69879. Preferably, Rep is added to host cells at an MOI of about 3. When the host cell line is a mammalian cell line, *e.g.*, HEK293 cells, the cell lines can have polynucleotide vector template stably integrated, and a second vector such as herpes virus can be used to introduce Rep protein into cells, allowing for the excision and amplification of ceDNA in the presence of Rep and helper virus.

[00152] In one embodiment, the host cells used to make the ceDNA vectors described herein are insect cells, and baculovirus is used to deliver both the polynucleotide that encodes a single Rep protein and the non-viral DNA vector polynucleotide expression construct template for ceDNA, *e.g.*, as described in FIGS. 4A-4C and Example 1. In some embodiments, the host cell is engineered to express a single Rep protein.

[00153] The ceDNA vector is then harvested and isolated from the host cells. The time for harvesting and collecting ceDNA vectors described herein from the cells can be selected and optimized to achieve a high-yield production of the ceDNA vectors. For example, the harvest time can be selected in view of cell viability, cell morphology, cell growth, etc. In one embodiment, cells are grown under sufficient conditions and harvested a sufficient time after baculoviral infection to produce ceDNA vectors but before a majority of cells start to die because of the baculoviral toxicity. The DNA vectors can be isolated using plasmid purification kits such as Qiagen Endo-Free Plasmid kits. Other methods developed for plasmid isolation can be also adapted for DNA vectors. Generally, any nucleic acid purification methods can be adopted.

[00154] The DNA vectors can be purified by any means known to those of skill in the art for purification of DNA. In one embodiment, ceDNA vectors are purified as DNA molecules. In another embodiment, the ceDNA vectors are purified as exosomes or microparticles.

[00155] The presence of the ceDNA vector can be confirmed by digesting the vector DNA isolated from the cells with a restriction enzyme having a single recognition site on the DNA vector and analyzing both digested and undigested DNA material using gel electrophoresis to confirm the presence of characteristic bands of linear and continuous DNA as compared to linear and non-continuous DNA. FIG. 4C and 4E illustrate one embodiment for identifying the presence of the closed ended ceDNA vectors produced by the processes herein. For example, FIG. 5 is a gel confirming the production of ceDNA from multiple plasmid constructs using one embodiment for producing these vectors as described in the Examples.

B. ceDNA Plasmid

[00156] A ceDNA-plasmid is a plasmid used for later production of a ceDNA vector. In some embodiments, a ceDNA-plasmid can be constructed using known techniques to provide at least the following as operatively linked components in the direction of transcription: (1) a 5' ITR sequence; (2) an expression cassette containing a cis-regulatory element, for example, a promoter, inducible promoter, regulatory switch, enhancers and the like; and (3) a 3' ITR sequence, where the 3' ITR sequence is asymmetric relative to the 5' ITR sequence. In some embodiments, the expression cassette flanked by the ITRs comprises a cloning site for introducing an exogenous sequence. The expression cassette replaces the rep and cap coding regions of the AAV genomes.

[00157] In one aspect, a ceDNA vector is obtained from a plasmid, referred to herein as a "ceDNA-plasmid" encoding in this order: a first adeno-associated virus (AAV) inverted terminal repeat (ITR), an expression cassette comprising a transgene, and a mutated or modified AAV ITR, wherein said ceDNA-plasmid is devoid of AAV capsid protein coding sequences. In alternative embodiments, the ceDNA-plasmid encodes in this order: a first (or 5') modified or mutated AAV ITR, an expression cassette comprising a transgene, and a second (or 3') wild-type AAV ITR, wherein said ceDNA-plasmid is devoid of AAV capsid protein coding sequences, and wherein the 5' and 3' ITRs are asymmetric relative to each other. In alternative embodiments, the ceDNA-plasmid encodes in this order: a first (or 5') modified or mutated AAV ITR, an expression cassette comprising a transgene, and a second (or 3') mutated or modified AAV ITR, wherein said ceDNA-plasmid is devoid of AAV capsid protein coding sequences, and wherein the 5' and 3' modified ITRs are different and do not have the same modifications.

[00158] In a further embodiment, the ceDNA-plasmid system is devoid of viral capsid protein coding sequences (*i.e.* it is devoid of AAV capsid genes but also of capsid genes of other viruses). In addition, in a particular embodiment, the ceDNA-plasmid is also devoid of AAV Rep protein coding sequences. Accordingly, in a preferred embodiment, ceDNA-plasmid is devoid of functional AAV cap and AAV rep genes GG-3′ for AAV2) plus a variable palindromic sequence allowing for hairpin formation.

[00159] A ceDNA-plasmid of the present invention can be generated using natural nucleotide sequences of the genomes of any AAV serotypes well known in the art. In one embodiment, the ceDNA-plasmid backbone is derived from the AAV1, AAV2, AAV3, AAV4, AAV5, AAV 5, AAV7, AAV8, AAV9, AAV10, AAV 11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 genome. *E.g.,* NCBI: NC 002077; NC 001401; NC001729; NC001829; NC006152; NC 006260; NC 006261; Kotin and Smith, The Springer Index of Viruses, available at the URL maintained by Springer (at www web address: oesys.springer.de/viruses/database/mkchapter.asp?virID=42.04.)(note -references to a URL or database refer to the contents of the URL or database as of the effective filing date of this application) In a particular embodiment, the ceDNA-plasmid backbone is derived from the AAV2 genome. In another particular embodiment, the ceDNA-plasmid backbone is a synthetic backbone genetically engineered to include at its 5' and 3' ITRs derived from one of these AAV genomes.

[00160] A ceDNA-plasmid can optionally include a selectable or selection marker for use in the establishment of a ceDNA vector-producing cell line. In one embodiment, the selection marker can be inserted downstream (*i.e.*, 3') of the 3' ITR sequence. In another embodiment, the selection marker can be inserted upstream (*i.e.*, 5') of the 5' ITR sequence. Appropriate selection markers include, for example, those that confer drug resistance. Selection markers can be, for example, a blasticidin S-resistance gene, kanamycin, geneticin, and the like. In a preferred embodiment, the drug selection marker is a blasticidin S-resistance gene.

[00161] An Exemplary ceDNA (*e.g.*, rAAV0) is produced from an rAAV plasmid. A method for the production of a rAAV vector, can comprise: (a) providing a host cell with a rAAV plasmid as described above, wherein both the host cell and the plasmid are devoid of capsid protein encoding genes, (b) culturing the host cell under conditions allowing production of an ceDNA genome, and (c) harvesting the cells and isolating the AAV genome produced from said cells.

C. Exemplary method of making the ceDNA vectors from ceDNA plasmids

[00162] Methods for making capsid-less ceDNA vectors are also provided herein, notably a method with a sufficiently high yield to provide sufficient vector for *in vivo* experiments.

[00163] In some embodiments, a method for the production of a ceDNA vector comprises the steps of: (1) introducing the nucleic acid construct comprising an expression cassette and two asymmetric ITR sequences into a host cell (*e.g.*, Sf9 cells), (2) optionally, establishing a clonal cell line, for example, by using a selection marker present on the plasmid, (3) introducing a Rep coding gene (either by transfection or infection with a baculovirus carrying said gene) into said insect cell, and (4) harvesting the cell and purifying the ceDNA vector. The nucleic acid construct comprising an expression cassette and two ITR sequences described above for the production of capsid-free AAV vector can be in the form of a cfAAV-plasmid, or Bacmid or Baculovirus generated with the cfAAV-plasmid as described below. The nucleic acid construct can be introduced into a host cell by transfection, viral transduction, stable integration, or other methods known in the art.

D. *Cell lines*:

[00164] Host cell lines used in the production of a ceDNA vector can include insect cell lines derived from Spodoptera frugiperda, such as Sf9, Sf21, or Trichoplusia ni cell, or other invertebrate, vertebrate, or other eukaryotic cell lines including mammalian cells. Other cell lines known to an ordinarily skilled artisan can also be used, such as HEK293, Huh-7, HeLa, HepG2, HeplA, 911, CHO, COS, MeWo, NIH3T3, A549, HT1080, monocytes, and mature and immature dendritic cells. Host cell lines can be transfected for stable expression of the ceDBA-plasmid for high yield ceDNA vector production.

[00165] ceDNA-plasmids can be introduced into Sf9 cells by transient transfection using reagents (*e.g.*, liposomal, calcium phosphate) or physical means (*e.g.*, electroporation) known in the art. Alternatively, stable Sf9 cell lines which have stably integrated the ceDNA-plasmid into their genomes can be established. Such stable cell lines can be established by incorporating a selection marker into the ceDNA -plasmid as described above. If the ceDNA -plasmid used to transfect the cell line includes a selection marker, such as an antibiotic, cells that have been transfected with the ceDNA-plasmid and integrated the ceDNA-plasmid DNA into their genome can be selected for by addition of the antibiotic to the cell growth media. Resistant clones of the cells can then be isolated by single-cell dilution or colony transfer techniques and propagated.

***E. Isolating and Purifying ceDNA vectors*:**

[00166] Examples of the process for obtaining and isolating ceDNA vectors are described in FIGS. 4A-4E and the specific examples below. ceDNA-vectors disclosed herein can be obtained from a producer cell expressing a single AAV Rep protein(s), further transformed with a ceDNA-plasmid, ceDNA-bacmid, or ceDNA-baculovirus. Plasmids useful for the production of ceDNA vectors include plasmids shown in FIGS. 8A (useful for Rep BIICs production), FIG. 8B (plasmid used to obtain a ceDNA vector).

[00167] In one aspect, a polynucleotide encodes the single AAV Rep protein (Rep 78 or 68) delivered to a producer cell in a plasmid (Rep-plasmid), a bacmid (Rep-bacmid), or a baculovirus (Rep-baculovirus). The Rep-plasmid, Rep-bacmid, and Rep-baculovirus can be generated by methods described above.

[00168] Methods to produce a ceDNA-vector, which is an exemplary ceDNA vector, are described herein. Expression constructs used for generating a ceDNA vectors of the present invention can be a plasmid (*e.g.*, ceDNA-plasmids), a Bacmid (*e.g.*, ceDNA-bacmid), and/or a baculovirus (*e.g.*, ceDNA-baculovirus). By way of an example only, a ceDNA-vector can be generated from the cells co-infected with ceDNA-baculovirus and Rep-baculovirus. Rep proteins produced from the Rep-baculovirus can replicate the ceDNA-baculovirus to generate ceDNA-vectors. Alternatively, ceDNA vectors can be generated from the cells stably transected with a construct comprising a sequence encoding a single AAV Rep protein (*e.g.*, Rep78, Rep68 or Rep52) delivered in Rep-plasmids, Rep-bacmids, or Rep-baculovirus. ceDNA-Baculovirus can be transiently transfected to the cells, be replicated by Rep protein and produce ceDNA vectors.

[00169] The bacmid (*e.g.*, ceDNA-bacmid) can be transfected into a permissive insect cells such as Sf9, Sf21, Tni (Trichoplusia ni) cell, High Five cell, and generate ceDNA-baculovirus, which is a recombinant

baculovirus including the sequences comprising the asymmetric ITRs and the expression cassette. ceDNA-baculovirus can be again infected into the insect cells to obtain a next generation of the recombinant baculovirus. Optionally, the step can be repeated once or multiple times to produce the recombinant baculovirus in a larger quantity.

[00170] The time for harvesting and collecting ceDNA vectors described herein from the cells can be selected and optimized to achieve a high-yield production of the ceDNA vectors. For example, the harvest time can be selected in view of cell viability, cell morphology, cell growth, etc. Usually, cells can be harvested after sufficient time after baculoviral infection to produce ceDNA vectors (*e.g.*, ceDNA vectors) but before majority of cells start to die because of the viral toxicity. The ceDNA-vectors can be isolated from the Sf9 cells using plasmid purification kits such as Qiagen ENDO-FREE PLASMID® kits. Other methods developed for plasmid isolation can be also adapted for ceDNA vectors. Generally, any art-known nucleic acid purification methods can be adopted, as well as commercially available DNA extraction kits.

[00171] Alternatively, purification can be implemented by subjecting a cell pellet to an alkaline lysis process, centrifuging the resulting lysate and performing chromatographic separation. As one nonlimiting example, the process can be performed by loading the supernatant on an ion exchange column (*e.g.* SARTOBIND Q®) which retains nucleic acids, and then eluting (*e.g.* with a 1.2 M NaCl solution) and performing a further chromatographic purification on a gel filtration column (*e.g.* 6 fast flow GE). The capsid-free AAV vector is then recovered by, *e.g.*, precipitation.

[00172] In some embodiments, ceDNA vectors can also be purified in the form of exosomes, or microparticles. It is known in the art that many cell types release not only soluble proteins, but also complex protein/nucleic acid cargoes *via* membrane microvesicle shedding (Cocucci *et al.*, 2009; EP 10306226.1). Such vesicles include microvesicles (also referred to as microparticles) and exosomes (also referred to as nanovesicles), both of which comprise proteins and RNA as cargo. Microvesicles are generated from the direct budding of the plasma membrane, and exosomes are released into the extracellular environment upon fusion of multivesicular endosomes with the plasma membrane. Thus, ceDNA vector-containing microvesicles and/or exosomes can be isolated from cells that have been transduced with the ceDNA-plasmid or a bacmid or baculovirus generated with the ceDNA-plasmid.

[00173] Microvesicles can be isolated by subjecting culture medium to filtration or ultracentrifugation at 20,000 x g, and exosomes at 100,000 x g. The optimal duration of ultracentrifugation can be experimentally-determined and will depend on the particular cell type from which the vesicles are isolated. Preferably, the culture medium is first cleared by low-speed centrifugation (*e.g.*, at 2000 x g for 5-20 minutes) and subjected to spin concentration using, *e.g.*, an AMICON® spin column (Millipore, Watford, UK). Microvesicles and exosomes can be further purified via FACS or MACS by using specific antibodies that recognize specific surface antigens present on the microvesicles and exosomes. Other microvesicle and exosome purification methods include, but are not limited to, immunoprecipitation, affinity chromatography, filtration, and magnetic beads coated with specific antibodies or aptamers. Upon

purification, vesicles are washed with, *e.g.*, phosphate-buffered saline. One advantage of using microvesicles or exosome to deliver ceDNA-containing vesicles is that these vesicles can be targeted to various cell types by including on their membranes proteins recognized by specific receptors on the respective cell types. (See also EP 10306226)

[00174] Another aspect of the invention herein relates to methods of purifying ceDNA vectors from host cell lines that have stably integrated a ceDNA construct into their own genome. In one embodiment, ceDNA vectors are purified as DNA molecules. In another embodiment, the ceDNA vectors are purified as exosomes or microparticles.

[00175] FIG. 5 shows a gel confirming the production of ceDNA from multiple ceDNA-plasmid constructs using the method described in the Examples. The ceDNA is confirmed by a characteristic band pattern in the gel, as discussed with respect to FIG. 4D in the Examples. Other characteristics of the ceDNA production process and intermediates are summarized in FIGS. 6A and 6B, and FIGS. 7A and 7B, as described in the Examples.

IV. ceDNA vector

[00176] As described herein, methods and compositions using a single Rep protein are useful in the production of a capsid-free ceDNA molecule with covalently-closed ends (ceDNA) vectors. In some embodiments, these ceDNA vectors can be produced in permissive host cells that comprises a single Rep protein, and are produced from an expression construct (*e.g.*, a ceDNA-plasmid, a ceDNA-bacmid, a ceDNA-baculovirus, or an integrated cell-line) containing a heterologous gene (transgene) positioned between two inverted terminal repeat (ITR) sequences, where the ITR sequences can be an asymmetrical ITR pair or a symmetrical- or substantially symmetrical ITR pair, as these terms are defined herein. A ceDNA vector comprising a NLS as disclosed herein can comprise ITR sequences that are selected from any of: (i) at least one WT ITR and at least one modified AAV inverted terminal repeat (mod-ITR) (*e.g.*, asymmetric modified ITRs); (ii) two modified ITRs where the mod-ITR pair have a different three-dimensional spatial organization with respect to each other (*e.g.*, asymmetric modified ITRs), or (iii) symmetrical or substantially symmetrical WT-WT ITR pair, where each WT-ITR has the same three-dimensional spatial organization, or (iv) symmetrical or substantially symmetrical modified ITR pair, where each mod-ITR has the same three-dimensional spatial organization, where the methods of the present disclosure may further include a delivery system, such as but not limited to a liposome nanoparticle delivery system.

[00177] The ceDNA vector is preferably duplex, *e.g.* self-complementary, over at least a portion of the molecule, such as the expression cassette (*e.g.* ceDNA is not a double stranded circular molecule). The ceDNA vector has covalently closed ends, and thus is resistant to exonuclease digestion (*e.g.* exonuclease I or exonuclease III), *e.g.* for over an hour at 37^0C.

[00178] A ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein has no packaging constraints imposed by the limiting space within the viral capsid.

ceDNA vectors represent a viable eukaryotically-produced alternative to prokaryote-produced plasmid DNA vectors, as opposed to encapsulated AAV genomes. This permits the insertion of control elements, *e.g.*, regulatory switches as disclosed herein, large transgenes, multiple transgenes etc.

[00179] In one aspect, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein comprises, in the 5' to 3' direction: a first adeno-associated virus (AAV) inverted terminal repeat (ITR), a nucleotide sequence of interest (for example an expression cassette as described herein) and a second AAV ITR, where the first ITR and the second ITR are asymmetric with respect to each other – that is, they are different from one another. As an exemplary embodiment, the first ITR can be a wild-type ITR and the second ITR can be a mutated or modified ITR. In some embodiments, the first ITR can be a mutated or modified ITR and the second ITR a wild-type ITR. In another embodiment, the first ITR and the second ITR are both modified but are different sequences, or have different modifications, or are not identical modified ITRs. Stated differently, the ITRs are asymmetric in that any changes in one ITR are not reflected in the other ITR; or alternatively, where the ITRs are different with respect to each other. Exemplary ITRs in the ceDNA vector and for use to generate a ceDNA-plasmid are discussed below in the section entitled "ITRs".

[00180] The wild-type or mutated or otherwise modified ITR sequences provided herein represent DNA sequences included in the expression construct (*e.g.*, ceDNA-plasmid, ce-DNA Bacmid, ceDNA-baculovirus) for production of the ceDNA vector. Thus, ITR sequences actually contained in the ceDNA vector produced from the ceDNA-plasmid or other expression construct may or may not be identical to the ITR sequences provided herein as a result of naturally occurring changes taking place during the production process (*e.g.*, replication error).

[00181] In some embodiments, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein comprises an expression cassette with a transgene, which can be, for example, a regulatory sequence, a sequence encoding a nucleic acid (*e.g.*, such as a miR or an antisense sequence), or a sequence encoding a polypeptide (*e.g.*, such as a transgene). In one embodiment, the transgene may be operatively linked to one or more regulatory sequence(s) that allows or controls expression of the transgene. In one embodiment, the polynucleotide comprises a first ITR sequence and a second ITR sequence, wherein the nucleotide sequence of interest is flanked by the first and second ITR sequences, and the first and second ITR sequences are asymmetrical relative to each other.

[00182] In one embodiment in each of these aspects, an expression cassette is located between two ITRs comprised in the following order with one or more of: a promoter operably linked to a transgene, a posttranscriptional regulatory element, and a polyadenylation and termination signal. In one embodiment, the promoter is regulatable - inducible or repressible. The promoter can be any sequence that facilitates the transcription of the transgene. In one embodiment the promoter is a CAG promoter (*e.g.* SEQ ID NO: 03), or variation thereof. The posttranscriptional regulatory element is a sequence that modulates expression of

the transgene, as a non-limiting example, any sequence that creates a tertiary structure that enhances expression of the transgene.

[00183] In one embodiment, the posttranscriptional regulatory element comprises WPRE (*e.g.* SEQ ID NO: 08). In one embodiment, the polyadenylation and termination signal comprises BGHpolyA (*e.g.* SEQ ID NO: 09). Any cis regulatory element known in the art, or combination thereof, can be additionally used *e.g.*, SV40 late polyA signal upstream enhancer sequence (USE), or other posttranscriptional processing elements including, but not limited to, the thymidine kinase gene of herpes simplex virus, or hepatitis B virus (HBV). In one embodiment, the expression cassette length in the 5' to 3' direction is greater than the maximum length known to be encapsidated in an AAV virion. In one embodiment, the length is greater than 4.6 kb, or greater than 5 kb, or greater than 6 kb, or greater than 7 kb. Various expression cassettes are exemplified herein.

[00184] An expression cassette in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise more than 4000 nucleotides, 5000 nucleotides, 10,000 nucleotides or 20,000 nucleotides, or 30,000 nucleotides, or 40,000 nucleotides or 50,000 nucleotides, or any range between about 4000-10,000 nucleotides or 10,000-50,000 nucleotides, or more than 50,000 nucleotides. In some embodiments, the expression cassette can comprise a transgene or nucleic acid in the range of 500 to 50,000 nucleotides in length. In some embodiments, the expression cassette can comprise a transgene or nucleic acid in the range of 500 to 75,000 nucleotides in length. In some embodiments, the expression cassette can comprise a transgene or nucleic acid is in the range of 500 to 10,000 nucleotides in length. In some embodiments, the expression cassette can comprise a transgene or nucleic acid is in the range of 1000 to 10,000 nucleotides in length. In some embodiments, the expression cassette can comprise a transgene or nucleic acid is in the range of 500 to 5,000 nucleotides in length. The ceDNA vectors do not have the size limitations of encapsidated AAV vectors, thus enable delivery of a large-size expression cassette to provide efficient expression of transgenes. In some embodiments, the ceDNA vector is devoid of prokaryote-specific methylation.

[00185] In some embodiments, the expression cassette in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can also comprise an internal ribosome entry site (IRES) and/or a 2A element. The cis-regulatory elements include, but are not limited to, a promoter, a riboswitch, an insulator, a mir-regulatable element, a post-transcriptional regulatory element, a tissue- and cell type-specific promoter and an enhancer. In some embodiments the ITR can act as the promoter for the transgene. In some embodiments, the ceDNA vector comprises additional components to regulate expression of the transgene, for example, one or more regulatory switches, which are described herein in the section entitled "Regulatory Switches" for controlling and regulating the expression of the transgene, and can include if desired, a regulatory switch which is a kill switch to enable controlled cell death of a cell comprising a ceDNA vector.

[00186] FIG. 1A-1C show schematics of nonlimiting, exemplary ceDNA vectors, or the corresponding sequence of ceDNA plasmids. ceDNA vectors are capsid-free and can be obtained from a plasmid encoding in this order: a first ITR, expressible transgene cassette and a second ITR, where at least one of the first and/or second ITR sequence is mutated with respect to the corresponding wild type AAV2 ITR sequence. The expressible transgene cassette preferably includes one or more of, in this order: an enhancer/promoter, an ORF reporter (transgene), a post-transcription regulatory element (*e.g.*, WPRE), and a polyadenylation and termination signal (*e.g.*, BGH polyA).

[00187] An expression cassette in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise any transgene of interest. Transgenes of interest include but are not limited to, nucleic acids encoding polypeptides, or non-coding nucleic acids (*e.g.*, RNAi, miRs etc.) preferably therapeutic (*e.g.*, for medical, diagnostic, or veterinary uses) or immunogenic (*e.g.*, for vaccines) polypeptides. In certain embodiments, the transgenes in the expression cassette encodes one or more polypeptides, peptides, ribozymes, peptide nucleic acids, siRNAs, RNAis, antisense oligonucleotides, antisense polynucleotides, antibodies, antigen binding fragments, or any combination thereof. In some embodiments, the transgene is a therapeutic gene, or a marker protein. In some embodiments, the transgene is an agonist or antagonist. In some embodiments, the antagonist is a mimetic or antibody, or antibody fragment, or antigen-binding fragment thereof, *e.g.*, a neutralizing antibody or antibody fragment and the like. In some embodiments, the transgene encodes an antibody, including a full-length antibody or antibody fragment, as defined herein. In some embodiments, the antibody is an antigen-binding domain or an immunoglobulin variable domain sequence, as that is defined herein.

[00188] In particular, the transgene can encode one or more therapeutic agent(s), including, but not limited to, for example, protein(s), polypeptide(s), peptide(s), enzyme(s), antibodies, antigen binding fragments, as well as variants, and/or active fragments thereof, for use in the treatment, prophylaxis, and/or amelioration of one or more symptoms of a disease, dysfunction, injury, and/or disorder. Exemplary transgenes are described herein in the section entitled “Method of Treatment”.

[00189] There are many structural features of ceDNA vectors produced according to the methods and compositions using a single Rep protein as disclosed herein that differ from plasmid-based expression vectors. ceDNA vectors may possess one or more of the following features: the lack of original (*i.e.* not inserted) bacterial DNA, the lack of a prokaryotic origin of replication, being self-containing, *i.e.*, they do not require any sequences other than the two ITRs, including the Rep binding and terminal resolution sites (RBS and TRS), and an exogenous sequence between the ITRs, the presence of ITR sequences that form hairpins, of the eukaryotic origin (*i.e.*, they are produced in eukaryotic cells), and the absence of bacterial-type DNA methylation or indeed any other methylation considered abnormal by a mammalian host. In general, it is preferred for the present vectors not to contain any prokaryotic DNA but it is contemplated that some prokaryotic DNA may be inserted as an exogenous sequence, as a nonlimiting example in a promoter or enhancer region. Another important feature distinguishing ceDNA vectors from plasmid expression

vectors is that ceDNA vectors are single-strand linear DNA having closed ends, while plasmids are always double-stranded DNA.

[00190] ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein preferably have a linear and continuous structure rather than a non-continuous structure, as determined by restriction enzyme digestion assay (FIG.4D). The linear and continuous structure is believed to be more stable from attack by cellular endonucleases, as well as less likely to be recombined and cause mutagenesis. Thus, a ceDNA vector in the linear and continuous structure is a preferred embodiment. The continuous, linear, single strand intramolecular duplex ceDNA vector can have covalently bound terminal ends, without sequences encoding AAV capsid proteins. These ceDNA vectors are structurally distinct from plasmids (including ceDNA plasmids described herein), which are circular duplex nucleic acid molecules of bacterial origin. The complimentary strands of plasmids may be separated following denaturation to produce two nucleic acid molecules, whereas in contrast, ceDNA vectors, while having complimentary strands, are a single DNA molecule and therefore even if denatured, remain a single molecule. In some embodiments, ceDNA vectors as described herein can be produced without DNA base methylation of prokaryotic type, unlike plasmids. Therefore, the ceDNA vectors and ceDNA-plasmids are different both in term of structure (in particular, linear versus circular) and also in view of the methods used for producing and purifying these different objects (see below), and also in view of their DNA methylation which is of prokaryotic type for ceDNA-plasmids and of eukaryotic type for the ceDNA vector.

[00191] Several advantages of a ceDNA vector described herein over plasmid-based expression vectors include, but are not limited to: 1) plasmids contain bacterial DNA sequences and are subjected to prokaryotic-specific methylation, *e.g.*, 6-methyl adenosine and 5-methyl cytosine methylation, whereas capsid-free AAV vector sequences are of eukaryotic origin and do not undergo prokaryotic-specific methylation; as a result, capsid-free AAV vectors are less likely to induce inflammatory and immune responses compared to plasmids; 2) while plasmids require the presence of a resistance gene during the production process, ceDNA vectors do not; 3) while a circular plasmid is not delivered to the nucleus upon introduction into a cell and requires overloading to bypass degradation by cellular nucleases, ceDNA vectors contain viral cis-elements, *i.e.*, ITRs, that confer resistance to nucleases and can be designed to be targeted and delivered to the nucleus. It is hypothesized that the minimal defining elements indispensable for ITR function are a Rep-binding site (RBS; 5'-GCGCGCTCGCTCGCTC-3' (SEQ ID NO: 531) for AAV2) and a terminal resolution site (TRS; 5'-AGTTGG-3' (SEQ ID NO: 48) for AAV2) plus a variable palindromic sequence allowing for hairpin formation; and 4) ceDNA vectors do not have the over-representation of CpG dinucleotides often found in prokaryote-derived plasmids that reportedly binds a member of the Toll-like family of receptors, eliciting a T cell-mediated immune response. In contrast, transductions with capsid-free AAV vectors disclosed herein can efficiently target cell and tissue-types that are difficult to transduce with conventional AAV virions using various delivery reagent.

V. ITRs

[00192] ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein comprise a heterologous gene positioned between two inverted terminal repeat (ITR) sequences, that differ with respect to each other (*i.e.* are asymmetric ITRs). In some embodiments, at least one of the ITRs is modified by deletion, insertion, and/or substitution as compared to a wild-type ITR sequence (*e.g.* AAV ITR); and at least one of the ITRs comprises a functional Rep binding site (RBS; *e.g.* 5'-GCGCGCTCGCTCGCTC-3' for AAV2, SEQ ID NO: 531) and a functional terminal resolution site (TRS; *e.g.* 5'-AGTT-3', SEQ ID NO: 46.) In one embodiment, at least one of the ITRs is a non-functional ITR. In one embodiment, the different ITRs are not each wild type ITRs from different serotypes.

[00193] While the ITRs exemplified in the specification and Examples herein are AAV2 ITRs, one of ordinary skill in the art is aware that one can as stated above use ITRs from any known parvovirus, for example a dependovirus such as AAV (*e.g.*, AAV1, AAV2, AAV3, AAV4, AAV5, AAV 5, AAV7, AAV8, AAV9, AAV10, AAV 11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 genome. *E.g.,* NCBI: NC 002077; NC 001401; NC001729; NC001829; NC006152; NC 006260; NC 006261), chimeric ITRs, or ITRs from any synthetic AAV. In some embodiments, the AAV can infect warm-blooded animals, *e.g.*, avian (AAAV), bovine (BAAV), canine, equine, and ovine adeno-associated viruses. In some embodiments the ITR is from B19 parvovirus (GenBank Accession No: NC 000883), Minute Virus from Mouse (MVM) (GenBank Accession No. NC 001510); goose parvovirus (GenBank Accession No. NC 001701); snake parvovirus 1 (GenBank Accession No. NC 006148).

[00194] In some embodiments, the ITR sequence in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can be from viruses of the Parvoviridae family, which includes two subfamilies: Parvovirinae, which infect vertebrates, and Densovirinae, which infect insects. The subfamily Parvovirinae (referred to as the parvoviruses) includes the genus Dependovirus, the members of which, under most conditions, require coinfection with a helper virus such as adenovirus or herpes virus for productive infection. The genus Dependovirus includes adeno-associated virus (AAV), which normally infects humans (*e.g.*, serotypes 2, 3A, 3B, 5, and 6) or primates (*e.g.*, serotypes 1 and 4), and related viruses that infect other warm-blooded animals (*e.g.*, bovine, canine, equine, and ovine adeno-associated viruses). The parvoviruses and other members of the Parvoviridae family are generally described in Kenneth I. Berns, "Parvoviridae: The Viruses and Their Replication," Chapter 69 in FIELDS VIROLOGY (3d Ed. 1996).

[00195] An ordinarily skilled artisan is aware that ITR sequences have a common structure of a double-stranded Holliday junction, which typically is a T-shaped or Y-shaped hairpin structure (see *e.g.*, FIG. 2A and FIG. 3A), where each ITR is formed by two palindromic arms or loops (B-B' and C-C') embedded in a larger palindromic arm (A-A'), and a single stranded D sequence, (where the order of these palindromic sequences defines the flip or flop orientation of the ITR), one can readily determine corresponding modified ITR sequences from any AAV serotype for use in a ceDNA vector or ceDNA-plasmid based on the

exemplary AAV2 ITR sequences provided herein. See, for example, structural analysis and sequence comparison of ITRs from different AAV serotypes (AAV1-AAV6) and described in Grimm *et al.*, J. Virology, 2006; 80(1); 426-439; Yan *et al.,* J. Virology, 2005; 364-379; Duan *et al.*, Virology 1999; 261; 8-14.

[00196] Specific alterations and mutations in the ITRs are described in detail herein, but in the context of ITRs, "altered" or "mutated" indicates that nucleotides have been inserted, deleted, and/or substituted relative to the wild-type, reference, or original ITR sequence, and can be altered relative to the other flanking ITR in a ceDNA vector having two flanking ITRs. The altered or mutated ITR can be an engineered ITR. As used herein, "engineered" refers to the aspect of having been manipulated by the hand of man. For example, a polypeptide is considered to be "engineered" when at least one aspect of the polypeptide, *e.g.,* its sequence, has been manipulated by the hand of man to differ from the aspect as it exists in nature.

[00197] In some embodiments, an ITR in ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein may be synthetic. In one embodiment, a synthetic ITR is based on ITR sequences from more than one AAV serotype. In another embodiment, a synthetic ITR includes no AAV-based sequence. In yet another embodiment, a synthetic ITR preserves the ITR structure described above although having only some or no AAV-sourced sequence. In some aspects, a synthetic ITR may interact preferentially with a wildtype Rep or a Rep of a specific serotype, or in some instances will not be recognized by a wild-type Rep and be recognized only by a mutated Rep.

[00198] ITR sequences have a common structure of a double-stranded Holliday junction, which typically is a T-shaped or Y-shaped hairpin structure (see, *e.g.*, FIG. 2A and FIG. 3A), where each ITR is formed by two palindromic arms or loops (B-B' and C-C') embedded in a larger palindromic arm (A-A'), and a single stranded D sequence, (where the order of these palindromic sequences defines the 'flip' or 'flop' orientation of the ITR). One of ordinary skill in the art can readily determine ITR sequences or modified ITR sequences from any AAV serotype for use in a ceDNA vector or ceDNA-plasmid based on the exemplary AAV2 ITR sequences provided herein. See, for example, the sequence comparison of ITRs from different AAV serotypes (AAV1-AAV6, and avian AAV (AAAV) and bovine AAV (BAAV)) described in Grimm *et al.*, J. Virology, 2006; 80(1); 426-439; that show the % identity of the left ITR of AAV2 to the left ITR from other serotypes: AAV-1 (84%), AAV-3 (86%), AAV-4 (79%), AAV-5 (58%), AAV-6 (left ITR) (100%) and AAV-6 (right ITR) (82%).

[00199] Accordingly, while the AAV2 ITRs are used as exemplary ITRs in ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein, a ceDNA vector may be prepared with or based on ITRs of any known AAV serotype, including, for example, AAV serotype 1 (AAV1), AAV serotype 2 (AAV2), AAV serotype 4 (AAV4), AAV serotype 5 (AAV5), AAV serotype 6 (AAV6), AAV serotype 7 (AAV7), AAV serotype 8 (AAV8), AAV serotype 9 (AAV9), AAV serotype 10 (AAV10), AAV serotype 11 (AAV11), or AAV serotype 12 (AAV12). The skilled artisan can

determine the corresponding sequence in other serotypes by known means. For example, determining if the change is in the A, A', B, B', C, C' or D region and determine the corresponding region in another serotype. One can use BLAST® (Basic Local Alignment Search Tool) or other homology alignment programs at default status to determine the corresponding sequence. The invention further provides populations and pluralities of ceDNA vectors comprising ITRs from a combination of different AAV serotypes – that is, one ITR can be from one AAV serotype and the other ITR can be from a different serotype. Without wishing to be bound by theory, in one embodiment one ITR can be from or based on an AAV2 ITR sequence and the other ITR of the ceDNA vector can be from or be based on any one or more ITR sequence of AAV serotype 1 (AAV1), AAV serotype 4 (AAV4), AAV serotype 5 (AAV5), AAV serotype 6 (AAV6), AAV serotype 7 (AAV7), AAV serotype 8 (AAV8), AAV serotype 9 (AAV9), AAV serotype 10 (AAV10), AAV serotype 11 (AAV11), or AAV serotype 12 (AAV12).

[00200] Any parvovirus ITR can be used as an ITR or as a base ITR for modification. Preferably, the parvovirus is a dependovirus. More preferably AAV. The serotype chosen can be based upon the tissue tropism of the serotype. AAV2 has a broad tissue tropism, AAV1 preferentially targets to neuronal and skeletal muscle, and AAV5 preferentially targets neuronal, retinal pigmented epithelia, and photoreceptors. AAV6 preferentially targets skeletal muscle and lung. AAV8 preferentially targets liver, skeletal muscle, heart, and pancreatic tissues. AAV9 preferentially targets liver, skeletal and lung tissue. In one embodiment, the modified ITR is based on an AAV2 ITR. For example, it is selected from the group consisting of: SEQ ID NO:2 and SEQ ID NO:52. In one embodiment of each of these aspects, the vector polynucleotide comprises a pair of ITRs, selected from the group consisting of: SEQ ID NO:1 and SEQ ID NO:52; and SEQ ID NO:2 and SEQ ID NO:51. In one embodiment of each of these aspects, the vector polynucleotide or the non-viral, capsid-free DNA vectors with covalently-closed ends comprises a pair of different ITRs selected from the group consisting of: SEQ ID NO:101 and SEQ ID NO:102; SEQ ID NO:103, and SEQ ID NO:104, SEQ ID NO:105, and SEQ ID NO:106; SEQ ID NO:107, and SEQ ID NO:108; SEQ ID NO:109, and SEQ ID NO:110; SEQ ID NO:111, and SEQ ID NO:112; SEQ ID NO:113 and SEQ ID NO:114; and SEQ ID NO:115 and SEQ ID NO:116. In some embodiments, a modified ITR is selected from any of the ITRs, or partial ITR sequences of SEQ ID NOS: 2, 52, 63, 64, 101-499 or 545-547.

[00201] In some embodiments, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise an ITR with a modification in the ITR corresponding to any of the modifications in ITR sequences or ITR partial sequences shown in any one or more of Tables 2, 3, 4, 5, 6, 7, 8, 9, 10A and 10B herein, or the sequences shown in FIG. 26A or 26B.

[00202] In some embodiments, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can form an intramolecular duplex secondary structure. The secondary structure of the first ITR and the asymmetric second ITR are exemplified in the context of wild-type ITRs (see, *e.g.*, FIGS. 2A, 3A, 3C) and modified ITR structures (see *e.g.*, FIGS. 2B and FIGS. 3B, 3D). Secondary structures are inferred or predicted based on the ITR sequences of the plasmid used to produce

the ceDNA vector. Exemplary secondary structures of the modified ITRs in which part of the stem-loop structure is deleted are shown in FIGS. 9A-25B and FIGS. 26A-26B, and also shown in Tables 10A and 10B. Exemplary secondary structures of the modified ITRs comprising a single stem and two loops are shown in FIGS. 9A-13B. Exemplary secondary structure of a modified ITR with a single stem and single loop is shown in FIG. 14. In some embodiments, the secondary structure can be inferred as shown herein using thermodynamic methods based on nearest neighbor rules that predict the stability of a structure as quantified by folding free energy change. For example, the structure can be predicted by finding the lowest free energy structure. In some embodiments, an algorithm disclosed in Reuter, J. S., & Mathews, D. H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics. 11,129 and implemented in the RNAstructure software (available at world wide web address: "rna.urmc.rochester.edu/RNAstructureWeb/index.html") can be used for prediction of the ITR structure. The algorithm can also include both free energy change parameters at 37°C and enthalpy change parameters derived from experimental literature to allow prediction of conformation stability at an arbitrary temperature. Using the RNA structure software, some of the modified ITR structures can be predicted as modified T-shaped stem-loop structures with estimated Gibbs free energy (ΔG) of unfolding under physiological conditions shown in FIGS. 3A-3D. Using the RNAstructure software, the three types of modified ITRs are predicted to have a Gibbs free energy of unfolding higher than a wild-type ITR of AAV2 (-92.9 kcal/mol) and are as follows: (a) The modified ITRs with a single-arm/single-unpaired-loop structure provided herein are predicted to have a Gibbs free energy of unfolding that ranges between -85 and -70 kcal/mol. (b) The modified ITRs with a single-hairpin structure provided herein are predicted to have a Gibbs free energy of unfolding that ranges between -70 and -40 kcal/mol. (c) The modified ITRs with a two-arm structure provided herein are predicted to have a Gibbs free energy of unfolding that ranges between -90 and -70 kcal/mol. Without wishing to be bound by a theory, the structures with higher Gibbs free energy are easier to be unfold for replication by Rep 68 or Rep 78 replication proteins. Thus, modified ITRs having higher Gibbs free energy of unfolding – *e.g.*, a single-arm/single-unpaired-loop structure, a single-hairpin structure, a truncated structure – tend to be replicated more efficiently than wild-type ITRs.

[00203] In one embodiment, the left ITR of a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein is modified or mutated with respect to a wild type (wt) AAV ITR structure, and the right ITR is a wild type AAV ITR. In one embodiment, the right ITR of the ceDNA vector is modified with respect to a wild type AAV ITR structure, and the left ITR is a wild type AAV ITR. In such an embodiment, a modification of the ITR (*e.g.*, the left or right ITR) can be generated by a deletion, an insertion, or substitution of one or more nucleotides from the wild type ITR derived from the AAV genome.

[00204] The ITRs used herein can be resolvable and non-resolvable, and selected for use in the ceDNA vectors are preferably AAV sequences, with serotypes 1, 2, 3, 4, 5, 6, 7, 8 and 9 being preferred. Resolvable AAV ITRs do not require a wild-type ITR sequence (*e.g.*, the endogenous or wild-type AAV ITR sequence may be altered by insertion, deletion, truncation and/or missense mutations), as long as the terminal repeat

mediates the desired functions, *e.g.*, replication, virus packaging, integration, and/or provirus rescue, and the like. Typically, but not necessarily, the ITRs are from the same AAV serotype, *e.g.*, both ITR sequences of the ceDNA vector are from AAV2. The ITRs may be synthetic sequences that function as AAV inverted terminal repeats, such as the "double-D sequence" as described in U.S. Pat. No. 5,478,745 to Samulski *et al.* While not necessary, the ITRs can be from the same parvovirus, *e.g.*, both ITR sequences are from AAV2.

[00205] In one embodiment, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can include an ITR structure that is mutated with respect to one of the wild type ITRs disclosed herein, but where the mutant or modified ITR still retains an operable Rep binding site (RBE or RBE') and terminal resolution site (*trs*). In one embodiment, the mutant ceDNA ITR includes a functional replication protein site (RPS-1) and a replication competent protein that binds the RPS-1 site is used in production.

[00206] In one embodiment, at least one of the ITRs in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein is a defective ITR with respect to Rep binding and/or Rep nicking. In one embodiment, the defect is at least 30% relative to a wild type reduction ITR, in other embodiments it is at least 35%..., 50%..., 65%..., 75%..., 85%..., 90%..., 95%..., 98%..., or completely lacking in function or any point in-between. The host cells do not express viral capsid proteins and the polynucleotide vector template is devoid of any viral capsid coding sequences. In one embodiment, the polynucleotide vector templates and host cells that are devoid of AAV capsid genes and the resultant protein also do not encode or express capsid genes of other viruses. In addition, in a particular embodiment, the nucleic acid molecule is also devoid of AAV Rep protein coding sequences

[00207] In some embodiments, the structural element of the ITR can be any structural element that is involved in the functional interaction of the ITR with a single large Rep protein (*e.g.*, Rep 78 or Rep 68). In certain embodiments, the structural element provides selectivity to the interaction of an ITR with a single large Rep protein, *i.e.*, determines at least in part which Rep protein functionally interacts with the ITR. In other embodiments, the structural element physically interacts with a single large Rep protein when the Rep protein is bound to the ITR. Each structural element can be, *e.g.*, a secondary structure of the ITR, a nucleotide sequence of the ITR, a spacing between two or more elements, or a combination of any of the above. In one embodiment, the structural elements are selected from the group consisting of an A and an A' arm, a B and a B' arm, a C and a C' arm, a D arm, a Rep binding site (RBE) and an RBE' (*i.e.*, complementary RBE sequence), and a terminal resolution sire (trs).

[00208] More specifically, the ability of a structural element of an ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein, to functionally interact with a particular single Rep protein, *e.g.*, large Rep protein or small Rep protein, can be altered by modifying the structural element. For example, the nucleotide sequence of the structural element can be modified as compared to the wild-type sequence of the ITR. In one embodiment, the structural element (*e.g.*, A arm, A' arm, B arm, B' arm, C arm, C' arm, D arm, RBE, RBE', and trs) of an ITR can be removed

and replaced with a wild-type structural element from a different parvovirus. For example, the replacement structure can be from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV13, snake parvovirus (*e.g.*, royal python parvovirus), bovine parvovirus, goat parvovirus, avian parvovirus, canine parvovirus, equine parvovirus, shrimp parvovirus, porcine parvovirus, or insect AAV. For example, the ITR can be an AAV2 ITR and the A or A' arm or RBE can be replaced with a structural element from AAV5. In another example, the ITR can be an AAV5 ITR and the C or C' arms, the RBE, and the trs can be replaced with a structural element from AAV2. In another example, the AAV ITR can be an AAV5 ITR with the B and B' arms replaced with the AAV2 ITR B and B' arms.

[00209] By way of example only, Table 1 indicates exemplary modifications of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) in regions of modified ITRs, where X is indicative of a modification of at least one nucleic acid (*e.g.*, a deletion, insertion and/ or substitution) in that section relative to the corresponding wild-type ITR. In some embodiments, any modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) in any of the regions of C and/or C' and/or B and/or B' retains three sequential T nucleotides (*i.e.*, TTT) in at least one terminal loop. For example, if the modification results in any of: a single arm ITR (*e.g.*, single C-C' arm, or a single B-B' arm), or a modified C-B' arm or C'-B arm, or a two arm ITR with at least one truncated arm (*e.g.*, a truncated C-C' arm and/or truncated B-B' arm), at least the single arm, or at least one of the arms of a two arm ITR (where one arm can be truncated) retains three sequential T nucleotides (*i.e.*, TTT) in at least one terminal loop. In some embodiments, a truncated C-C' arm and/or a truncated B-B' arm has three sequential T nucleotides (*i.e.*, TTT) in the terminal loop.

Table 1: Exemplary combinations of modifications of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) to different B-B' and C-C' regions or arms of ITRs (X indicates a nucleotide modification, *e.g.*, addition, deletion or substitution of at least one nucleotide in the region).

B region	B' region	C region	C' region
X			
	X		
X	X		
		X	
			X
		X	X
X		X	
X			X

	X	X	
	X		X
X	X	X	
X	X		X
X		X	X
	X	X	X
X	X	X	X

[00210] In some embodiments, a modified ITR for use in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise any one of the combinations of modifications shown in Table 1, and also a modification of at least one nucleotide in any one or more of the regions selected from: between A' and C, between C and C', between C' and B, between B and B' and between B' and A. In some embodiments, any modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) in the C or C' or B or B' regions, still preserves the terminal loop of the stem-loop. In some embodiments, any modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) between C and C' and/or B and B' retains three sequential T nucleotide (*i.e.*, TTT) in at least one terminal loop. In alternative embodiments, any modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) between C and C' and/or B and B' retains three sequential "A" nucleotides (*i.e.*, AAA) in at least one terminal loop. In some embodiments, a modified ITR for use in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise any one of the combinations of modifications shown in Table 1, and also a modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) in any one or more of the regions selected from: A', A and/or D. For example, in some embodiments, a modified ITR for use herein can comprise any one of the combinations of modifications shown in Table 1, and also a modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) in the A region. In some embodiments, a modified ITR for use in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise any one of the combinations of modifications shown in Table 1, and also a modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) in the A' region. In some embodiments, a modified ITR for use herein can comprise any one of the combinations of modifications shown in Table 1, and also a modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) in the A and/or A' region. In some embodiments, a modified ITR for use herein can comprise any one of the combinations of modifications shown in Table 1, and also a modification of at least one nucleotide (*e.g.*, a deletion, insertion and/ or substitution) in the D region.

[00211] In one embodiment, the nucleotide sequence of the structural element of an ITR can be modified (*e.g.*, by modifying 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more nucleotides or any range therein) to produce a modified structural element. In one embodiment, the specific modifications to the ITRs in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein are exemplified herein (*e.g.*, SEQ ID NOS: 2, 52, 63, 64, 101-499, or 545-547). In some embodiments, an ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can be modified (*e.g.*, by modifying 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more nucleotides or any range therein). In other embodiments, an ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein ,can have at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or more sequence identity with one of the modified ITRs of SEQ ID NOS: 469-499 or 545-547, or the RBE-containing section of the A-A' arm and C-C' and B-B' arms of SEQ ID NO: 101-134 or 545-547.

[00212] In some embodiments, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can, for example, comprise removal or deletion of all of a particular arm, *e.g.*, all or part of the A-A' arm, or all or part of the B-B' arm or all or part of the C-C' arm, or alternatively, the removal of 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs forming the stem of the loop so long as the final loop capping the stem (*e.g.*, single arm) is still present (*e.g.*, see ITR-6). In some embodiments, a modified ITR can comprise the removal of 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs from the B-B' arm. In some embodiments, a modified ITR can comprise the removal of 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs from the C-C' arm. In some embodiments, a modified ITR can comprise the removal of 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs from the C-C' arm and the removal of 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs from the B-B' arm. Any combination of removal of base pairs is envisioned, for example, 6 base pairs can be removed in the C-C' arm and 2 base pairs in the B-B' arm. As an illustrative example, FIG. 13A-13B show an exemplary modified ITR with at least 7 base pairs deleted from each of the C portion and the C' portion, a substitution of a nucleotide in the loop between C and C' region, and at least one base pair deletion from each of the B region and B' regions such that the modified ITR comprises two arms where at least one arm (*e.g.*, C-C') is truncated. Note in this example, as the modified ITR comprises at least one base pair deletion from each of the B region and B' regions, arm B-B' is also truncated relative to WT ITR.

[00213] In some embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9 or more complementary base pairs are removed from each of the C portion and the C' portion of the C-C' arm such that the C-C' arm is truncated. That is, if a base is removed in the C portion of the C-C' arm, the complementary base pair in the C' portion is removed, thereby truncating the C-C' arm. In such embodiments, 2, 4, 6, 8 or more base pairs are removed from the C-C' arm such that the C-C' arm is truncated. In alternative embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs are removed from the C portion of the C-C' arm such that only C' portion of the arm remains. In

alternative embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs are removed from the C' portion of the C-C' arm such that only C portion of the arm remains.

[00214] In some embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9 or more complementary base pairs are removed from each of the B portion and the B' portion of the B-B' arm such that the B-B' arm is truncated. That is, if a base is removed in the B portion of the B-B' arm, the complementary base pair in the B' portion is removed, thereby truncating the B-B' arm. In such embodiments, 2, 4, 6, 8 or more base pairs are removed from the B-B' arm such that the B-B' arm is truncated. In alternative embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs are removed from the B portion of the B-B' arm such that only B' portion of the arm remains. In alternative embodiments, 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs are removed from the B' portion of the B-B' arm such that only B portion of the arm remains.

[00215] In some embodiments, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein, can have between 1 and 50 (*e.g.* 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50) nucleotide deletions relative to a full-length wild-type ITR sequence. In some embodiments, a modified ITR can have between 1 and 30 nucleotide deletions relative to a full-length WT ITR sequence. In some embodiments, a modified ITR has between 2 and 20 nucleotide deletions relative to a full-length wild-type ITR sequence.

[00216] In some embodiments, a modified ITR forms two opposing, lengthwise-asymmetric stem-loops, *e.g.*, C-C' loop is a different length to the B-B' loop. In some embodiments, one of the opposing, lengthwise-asymmetric stem-loops of a modified ITR has a C-C' and/or B-B' stem portion in the range of 8 to 10 base pairs in length and a loop portion (*e.g.*, between C-C' or between B-B') having 2 to 5 unpaired deoxyribonucleotides. In some embodiments, a one lengthwise-asymmetric stem-loop of a modified ITR has a C-C' and/or B-B' stem portion of less than 8, or less than 7, 6, 5, 4, 3, 2, 1 base pairs in length and a loop portion (*e.g.*, between C-C' or between B-B') having between 0-5 nucleotides. In some embodiments, a modified ITR with a lengthwise-asymmetric stem-loop has a C-C' and/or B-B' stem portion less than 3 base pairs in length.

[00217] In some embodiments, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein does not contain any nucleotide deletions in the RBE-containing portion of the A or A' regions, so as not to interfere with DNA replication (*e.g.* binding to a RBE by Rep protein, or nicking at a terminal resolution site). In some embodiments, a modified ITR encompassed for use in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein has one or more deletions in the B, B', C, and/or C' region as described herein. Several non-limiting examples of modified ITRs are shown in FIGS 9A-26B.

[00218] In some embodiments, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise a deletion of the B-B' arm, so that the C-C' arm remains, for example, see exemplary ITR-2 (left) and ITR-2 (right) shown in FIG 9A-9B and

ITR-4 (left) and ITR-4 (right) (FIGS, 11A-11B). In some embodiments, a modified ITR can comprise a deletion of the C-C' arm such that the B-B' arm remains, for example, see exemplary ITR-3 (left) and ITR-3 (right) shown in FIG 10A-10B. In some embodiments, a modified ITR can comprise a deletion of the B-B' arm and C-C' arm such that a single stem-loop remains, for example, see exemplary ITR-6 (left) and ITR-6 (right) shown in FIG 14A-14B, and ITR-21 and ITR-37. In some embodiments, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise a deletion of the C' region such that a truncated C-loop and B-B' arm remains, for example, see exemplary ITR-1 (left) and ITR-1 (right) shown in FIG 15A-15B. Similarly, in some embodiments, a modified ITR can comprise a deletion of the C region such that a truncated C'-loop and B-B' arm remains, for example, see exemplary ITR-5 (left) and ITR-5 (right) shown in FIG 16A-16B.

[00219] In some embodiments, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise a deletion of base pairs in any one or more of: the C portion, the C' portion, the B portion or the B' portion, such that complementary base pairing occurs between the C-B' portions and the C'-B portions to produce a single arm, for example, see ITR-10 (right) and ITR-10 (left) (FIG 12A-12B).

[00220] In some embodiments, in addition to a modification in one or more nucleotides in the C, C', B and/or B' regions, a modified ITR for use in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can comprise a modification (*e.g.*, deletion, substitution or addition) of at least 1, 2, 3, 4, 5, 6 nucleotides in any one or more of the regions selected from: between A' and C, between C and C', between C' and B, between B and B' and between B' and A. For example, the nucleotide between B' and C in a modified right ITR can be substituted from an A to a G, C or A or deleted or one or more nucleotides added; a nucleotide between C' and B in a modified left ITR can be changed from a T to a G, C or A, or deleted or one or more nucleotides added.

[00221] In certain embodiments, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein does not have a modified ITR consisting of the nucleotide sequence selected from any of: SEQ ID NOs: 550-557. In certain embodiments, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein does not have a modified ITR comprising the nucleotide sequence selected from any of: SEQ ID NOs: 550-557.

[00222] In some embodiments, the ceDNA vector comprises a regulatory switch as disclosed herein and a modified ITR selected having the nucleotide sequence selected from any of the group consisting of: SEQ ID NO: 550-557.

[00223] In another embodiment, the structure of the structural element of an ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can be modified. For example, the structural element a change in the height of the stem and/or the number of nucleotides in the loop. For example, the height of the stem can be about 2, 3, 4, 5, 6, 7, 8, or 9 nucleotides or more or any range therein. In one embodiment, the stem height can be about 5 nucleotides to about 9

nucleotides and functionally interacts with Rep. In another embodiment, the stem height can be about 7 nucleotides and functionally interacts with Rep. In another example, the loop can have 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides or more or any range therein.

[00224] In another embodiment, the number of GAGY binding sites or GAGY-related binding sites within the RBE or extended RBE can be increased or decreased. In one example, the RBE or extended RBE, can comprise 1, 2, 3, 4, 5, or 6 or more GAGY binding sites or any range therein. Each GAGY binding site can independently be an exact GAGY sequence or a sequence similar to GAGY as long as the sequence is sufficient to bind a Rep protein.

[00225] In another embodiment, the spacing between two elements (such as but not limited to the RBE and a hairpin) can be altered (*e.g.*, increased or decreased) to alter functional interaction with a single large Rep protein. For example, the spacing can be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 21 nucleotides or more or any range therein.

[00226] a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein described herein can include an ITR structure that is modified with respect to the wild type AAV2 ITR structure disclosed herein, but still retains an operable RBE, trs and RBE´ portion. FIG. 2A and FIG. 2B show one possible mechanism for the operation of a trs site within a wild type ITR structure portion of a ceDNA vector. In some embodiments, the ceDNA vector contains one or more functional ITR polynucleotide sequences that comprise a Rep-binding site (RBS; 5'-GCGCGCTCGCTCGCTC-3' (SEQ ID NO: 531) for AAV2) and a terminal resolution site (TRS; 5'-AGTT (SEQ ID NO: 46)). In some embodiments, at least one ITR (wt or modified ITR) is functional. In alternative embodiments, where a ceDNA vector comprises two modified ITRs that are different or asymmetrical to each other, at least one modified ITR is functional and at least one modified ITR is non-functional.

[00227] In some embodiments, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein does not have a modified ITR selected from any sequence consisting of, or consisting essentially of: SEQ ID NOs:500-529, as provided herein. In some embodiments, a ceDNA vector does not have an ITR that is selected from any sequence selected from SEQ ID NOs: 500-529.

[00228] In some embodiments, the modified ITR (*e.g.*, the left or right ITR) of a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein has modifications within the loop arm, the truncated arm, or the spacer. Exemplary sequences of ITRs having modifications within the loop arm, the truncated arm, or the spacer are listed in Table 2.

[00229] In some embodiments, the modified ITR (*e.g.*, the left or right ITR) of a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein has modifications within the loop arm and the truncated arm. Exemplary sequences of ITRs having modifications within the loop arm and the truncated arm are listed in Table 3.

[00230] In some embodiments, the modified ITR (*e.g.*, the left or right ITR) of a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein has modifications within the loop arm and the spacer. Exemplary sequences of ITRs having modifications within the loop arm and the spacer are listed in Table 4.

[00231] In some embodiments, the modified ITR (*e.g.*, the left or right ITR) of a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein has modifications within the truncated arm and the spacer. Exemplary sequences of ITRs having modifications within the truncated arm and the spacer are listed in Table 5.

[00232] In some embodiments, the modified ITR (*e.g.*, the left or right ITR) of a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein has modifications within the loop arm, the truncated arm, and the spacer. Exemplary sequences of ITRs having modifications within the loop arm, the truncated arm, and the spacer are listed in Table 6.

[00233] In some embodiments, an ITR (*e.g.*, the left or right ITR) in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein is modified such that it comprises the lowest energy of unfolding ("low energy structure"). A low energy will have reduced Gibbs free energy as compared to a wild type ITR. Exemplary sequences of ITRs that are modified to low (*i.e.*, reduced) energy of unfolding are presented herein in Table 7-9.

[00234] In some embodiments, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein is selected from any or a combination of those shown in Table 2-9, 10A or 10B.

[00235] Table 2: ITR Sequences with Modifications in Loop Arm, Truncated Arm or Spacer. These include the RBS sequence GCGCGCTCGCTCGCTC (SEQ ID NO: 531) at the 5' end and the complementary RBE' sequence GAGCGAGCGAGCGCGC (SEQ ID NO: 536) on the most 3' end.

Table 2

SEQ ID	Modified Region	Sequence	ΔG	No. Strut.
135	Truncated Arm	GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**A**GCCCGGGC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-73.6	1
136		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGAC**A**CCCGGG**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-73.7	3
137		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGACG**A**CCGG**T**CGGCCTCAGTGAGCGAGCGAGCGCGC	-74.2	1
138		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGACGC**A**CG**T**GCGGCCTCAGTGAGCGAGCGAGCGCGC	-75.7	2
139		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGACGCC**AT**GGCGGCCTCAGTGAGCGAGCGAGCGCGC	-75.2	1
140		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**A**G**A**CCGG**T**C**T**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
141		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGAC**A**C**A**CG**T**G**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-74.2	1

142		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGACG**A**C**AT**G**T**CGGCCTCAGTGAGCGAGCGAGCGCGC	-73.3	2
143		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**A**GC**A**CG**T**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-74.1	1
144		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AA**CCCGGG**TT**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
145		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**A**GCC**AT**GGC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-73.6	1
146		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGAC**AA**CCGG**TT**GGCCTCAGTGAGCGAGCGAGCGCGC	-73.0	1
147		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGAC**A**CC**AT**GG**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-73.7	1
148		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGACG**AA**CG**TT**CGGCCTCAGTGAGCGAGCGAGCGCGC	-73.7	1
149		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGACGC**AATT**GCGGCCTCAGTGAGCGAGCGAGCGCGC	-75.7	1
150		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AAA**CCGG**TTT**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
151		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AG AA**CG**TT**C**T**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
152		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**A**GC**AATT**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-74.1	1
153		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGAC**AAA**CG**TTT**GGCCTCAGTGAGCGAGCGAGCGCGC	-73.0	1
154		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGACG**AAATTT**CGGCCTCAGTGAGCGAGCGAGCGCGC	-73.7	2
155		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**A**GC**AATT**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-74.1	1
156		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AA**CC**AT**GG**TT**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
157		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGAC**AA**C**AT**G**TT**GGCCTCAGTGAGCGAGCGAGCGCGC	-73.0	1
158		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AG AC AT**G**T**C**T**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	2
159		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGAC**A**C**AATT**G**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-74.2	1
160		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AAAA**CG**TTTT**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
161		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AAA**C**AT**G**TTT**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	2
162		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AA**C**AATT**G**TT**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
163		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AG AAATTT**C**T**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
164		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGAC**AAAATTTT**GGCCTCAGTGAGCGAGCGAGCGCGC	-73.0	1
165		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCC CGA**AAAAATTTTT**GCCTCAGTGAGCGAGCGAGCGCGC	-72.1	1
166	Spacer	GCGCGCTCGCTCGCTC**G**CTGAGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCTCAG**C**GAGCGAGCGAGCGCGC	-76.7	1

167		GCGCGCTCGCTCGCTCA**A**TGAGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCTCA**T**TGAGCGAGCGAGCGCGC	-72.9	1
168		GCGCGCTCGCTCGCTCAC**C**GAGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCTC**G**GTGAGCGAGCGAGCGCGC	-76.7	1
169		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-72.9	1
170		GCGCGCTCGCTCGCTCACTG**G**GGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCC**C**CAGTGAGCGAGCGAGCGCGC	-77.3	2
171		GCGCGCTCGCTCGCTCACTGA**A**GCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGC**T**TCAGTGAGCGAGCGAGCGCGC	-72.8	1
172		GCGCGCTCGCTCGCTCACTGAG**A**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**T**CTCAGTGAGCGAGCGAGCGCGC	-73.1	1
173		GCGCGCTCGCTCGCTC**GA**TGAGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCTCA**TC**GAGCGAGCGAGCGCGC	-74.7	1
174		GCGCGCTCGCTCGCTC**G**C**G**GAGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCTC**C**G**C**GAGCGAGCGAGCGCGC	-78.2	2
175		GCGCGCTCGCTCGCTC**G**CT**A**AGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-72.5	1
176		GCGCGCTCGCTCGCTC**G**CTG**G**GGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCC**C**CAG**C**GAGCGAGCGAGCGCGC	-78.8	2
177		GCGCGCTCGCTCGCTC**G**CTGA**A**GCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGC**T**TCAG**C**GAGCGAGCGAGCGCGC	-74.3	1
178		GCGCGCTCGCTCGCTC**G**CTGAG**A**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**T**CTCAG**C**GAGCGAGCGAGCGCGC	-74.6	1
179		GCGCGCTCGCTCGCTC**GAG**GAGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCTC**CTC**GAGCGAGCGAGCGCGC	-76.9	1
180		GCGCGCTCGCTCGCTC**GA**TA**A**GGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCT**T**A**TC**GAGCGAGCGAGCGCGC	-72.4	1
181		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-73.8	2
182		GCGCGCTCGCTCGCTC**GA**TGA**A**GCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGC**T**TCA**TC**GAGCGAGCGAGCGCGC	-72.3	1
183		GCGCGCTCGCTCGCTC**GA**TGAG**A**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**T**CTCA**TC**GAGCGAGCGAGCGCGC	-72.6	1
184		GCGCGCTCGCTCGCTC**GAGA**AGGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCT**TCTC**GAGCGAGCGAGCGCGC	-74.5	1
185		GCGCGCTCGCTCGCTC**GAG**G**G**GGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCC**C**C**CTC**GAGCGAGCGAGCGCGC	-79	2
186		GCGCGCTCGCTCGCTC**GAG**GA**A**GCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGC**T**TC**CTC**GAGCGAGCGAGCGCGC	-74.5	1
189		GCGCGCTCGCTCGCTC**GAG**GAG**A**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**T**CTC**CTC**GAGCGAGCGAGCGCGC	-74.8	1
187		GCGCGCTCGCTCGCTC**GAG**G**G**GGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCC**C**C**CTC**GAGCGAGCGAGCGCGC	-79	2
188		GCGCGCTCGCTCGCTC**GAG**GA**A**GCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGC**T**TC**CTC**GAGCGAGCGAGCGCGC	-74.5	1
189		GCGCGCTCGCTCGCTC**GAG**GAG**A**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**T**CTC**CTC**GAGCGAGCGAGCGCGC	-74.8	1
190		GCGCGCTCGCTCGCTC**GAGAG**GGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCC**CTCTC**GAGCGAGCGAGCGCGC	-76.9	2

200		GCGCGCTCGCTCGCTC**GAGA**A**A**GCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGC**TT****TCTC**GAGCGAGCGAGCGCGC	-72.1	1
201		GCGCGCTCGCTCGCTC**GAGA**AG**A**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCCT**TCTC**GAGCGAGCGAGCGCGC	-69.1	2
202		GCGCGCTCGCTCGCTC**GAGAGA**GCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGC**TCTCTC**GAGCGAGCGAGCGCGC	-74.8	1
203		GCGCGCTCGCTCGCTC**GAGAG**G**A**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**T**C**CTCTC**GAGCGAGCGAGCGCGC	-74.8	1
204		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-72.4	1
205		GCGCGCTCGCTCGCTCA**AGAGAA**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**TTCTCT**TGAGCGAGCGAGCGCGC	-70.6	1
206		GCGCGCTCGCTCGCTCAC**GAGAA**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**TTCTC**GTGAGCGAGCGAGCGCGC	-72.2	1
207		GCGCGCTCGCTCGCTCACT**AGAA**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**TTCT**AGTGAGCGAGCGAGCGCGC	-70.8	1
208		GCGCGCTCGCTCGCTCACTG**GAA**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**TTC**CAGTGAGCGAGCGAGCGCGC	-72.8	1
209		GCGCGCTCGCTCGCTCACTGA**AA**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**TT**TCAGTGAGCGAGCGAGCGCGC	-70.4	1
210		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-80.3	2
211		GCGCGCTCGCTCGCTCA**A**T**A**A**AA**CCGGGCGACCAAAGGTCGCC CGACGCCCGGGCGG**TT**T**T**AT**T**GAGCGAGCGAGCGCGC	-65.8	1
212	Loop Arm	GCGCGCTCGCTCGCTCACTGAGGCC**A**GGCGACCAAAGGTCGCC **T**GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.7	1
213		GCGCGCTCGCTCGCTCACTGAGGCCG**A**GCGACCAAAGGTCGC**T** CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.1	1
214		GCGCGCTCGCTCGCTCACTGAGGCCGG**A**CGACCAAAGGTCG**T**C CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.1	2
215		GCGCGCTCGCTCGCTCACTGAGGCCGGG**A**GACCAAAGGTC**T**CC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.9	1
216		GCGCGCTCGCTCGCTCACTGAGGCCGGGC**A**ACCAAAGGT**T**GCC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.4	1
217		GCGCGCTCGCTCGCTCACTGAGGCCGGGCG**G**CCAAAGG**C**CGCC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-77.3	2
218		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGA**A**CAAAG**T**TCGCC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-72.8	2
219		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGAC**A**AAA**T**GTCGCC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.5	1
220		GCGCGCTCGCTCGCTCACTGAGGCC**AA**GCGACCAAAGGTCGC**T** **T**GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-71.3	1
221		GCGCGCTCGCTCGCTCACTGAGGCC**AAA**CGACCAAAGGTCG**TTT** GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-68.9	1
222		GCGCGCTCGCTCGCTCACTGAGGCC**AAAA**GACCAAAGGTC**TTTT** GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-67.3	2
223		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAA**ACCAAAGGT**TTTTT** GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-64.6	2
224		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAG**CCAAAGG**CTTTTT** GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-67	2

225		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAGA**CAAAG**TCTTTTT** GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-64.9	1
226		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAGAA**AAA**TTCTTTTT** GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-63.1	1
227		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAAAA**AAA**TTTTTTTT** GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-60.4	1
228		GCGCGCTCGCTCGCTCACTGAGGCCG**AAAAAAA**AAA**TTTTTTT**C GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-62.2	1
229		GCGCGCTCGCTCGCTCACTGAGGCCGG**AAAGAA**AAA**TTCTTT**CC GACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-67.3	1
230		GCGCGCTCGCTCGCTCACTGAGGCCGGG**AAGAA**AAA**TTCTT**CC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-69.7	2
231		GCGCGCTCGCTCGCTCACTGAGGCCGGGC**AGAA**AAA**TTCT**GCC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-71.9	1
232		GCGCGCTCGCTCGCTCACTGAGGCCGGGCG**GAA**AAA**TTC**CGCC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.4	2
233		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGA**AA**AAA**TT**TCGCC CGACGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC	-71.0	2

Table 3: modified ITR Sequences with Modifications in Loop Arm and Truncated Arm

Table 3				
SEQ ID	**Modified Region**	**Sequence**	**ΔG**	**No. Strut.**
234	Loop Arm & Truncated Arm	GCGCGCTCGCTCGCTCACTGAGGCC**A**GGCGACCAAAGGTCGCC**T**GA C**A**CCCGGG**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-72.2	2
235		GCGCGCTCGCTCGCTCACTGAGGCC**A**GGCGACCAAAGGTCGCC**T**GA CGCC**AT**GGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.7	1
236		GCGCGCTCGCTCGCTCACTGAGGCC**A**GGCGACCAAAGGTCGCC**T**GA CG**A**C**AT**G**T**CGGCCTCAGTGAGCGAGCGAGCGCGC	-71.8	1
237		GCGCGCTCGCTCGCTCACTGAGGCC**A**GGCGACCAAAGGTCGCC**T**GA CG**AA**CG**TT**CGGCCTCAGTGAGCGAGCGAGCGCGC	-72.2	1
238		GCGCGCTCGCTCGCTCACTGAGGCC**A**GGCGACCAAAGGTCGCC**T**GA **A**GC**AATT**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-72.6	1
239		GCGCGCTCGCTCGCTCACTGAGGCCGGGCG**G**CCAAAGG**C**CGCCCGA C**A**CCCGGG**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-75.8	2
240		GCGCGCTCGCTCGCTCACTGAGGCCGGGCG**G**CCAAAGG**C**CGCCCGA CGCC**AT**GGCGGCCTCAGTGAGCGAGCGAGCGCGC	-77.3	1
241		GCGCGCTCGCTCGCTCACTGAGGCCGGGCG**G**CCAAAGG**C**CGCCCGA CG**A**C**AT**G**T**CGGCCTCAGTGAGCGAGCGAGCGCGC	-75.4	1
242		GCGCGCTCGCTCGCTCACTGAGGCCGGGCG**G**CCAAAGG**C**CGCCCGA CG**AA**CG**TT**CGGCCTCAGTGAGCGAGCGAGCGCGC	-75.8	1
243		GCGCGCTCGCTCGCTCACTGAGGCCGGGCG**G**CCAAAGG**C**CGCCCGA **A**GC**AATT**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-76.2	1
244		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGAC**A**AAA**T**GTCGCCCGA C**A**CCCGGG**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-72	1
245		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGAC**A**AAA**T**GTCGCCCGA CGCC**AT**GGCGGCCTCAGTGAGCGAGCGAGCGCGC	-73.5	1
246		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGAC**A**AAA**T**GTCGCCCGA	-71.6	2

		CG**A**C**AT**G**T**CGGCCTCAGTGAGCGAGCGAGCGCGC		
247		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGAC**A**AAA**T**GTCGCCCGA CG**AA**CG**TT**CGGCCTCAGTGAGCGAGCGAGCGCGC	-72	2
248		GCGCGCTCGCTCGCTCACTGAGGCCGGGCGAC**A**AAA**T**GTCGCCCGA **A**GC**AATT**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-72.4	1
249		GCGCGCTCGCTCGCTCACTGAGGCC**AAAA**GACCAAAGGTC**TTTT**GA C**A**CCCGGG**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-65.8	3
250		GCGCGCTCGCTCGCTCACTGAGGCC**AAAA**GACCAAAGGTC**TTTT**GA CGCC**AT**GGCGGCCTCAGTGAGCGAGCGAGCGCGC	-67.3	2
251		GCGCGCTCGCTCGCTCACTGAGGCC**AAAA**GACCAAAGGTC**TTTT**GA CG**A**C**AT**G**T**CGGCCTCAGTGAGCGAGCGAGCGCGC	-65.4	2
252		GCGCGCTCGCTCGCTCACTGAGGCC**AAAA**GACCAAAGGTC**TTTT**GA CG**AA**CG**TT**CGGCCTCAGTGAGCGAGCGAGCGCGC	-65.8	2
253		GCGCGCTCGCTCGCTCACTGAGGCC**AAAA**GACCAAAGGTC**TTTT**GA **A**GC**AATT**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-66.2	1
254		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAAAA**AAA**TTTTTTTT**GA C**A**CCCGGG**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-59.6	2
255		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAAAA**AAA**TTTTTTTT**GA CGCC**AT**GGCGGCCTCAGTGAGCGAGCGAGCGCGC	-60.4	1
256		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAAAA**AAA**TTTTTTTT**GA CG**A**C**AT**G**T**CGGCCTCAGTGAGCGAGCGAGCGCGC	-59.8	1
257		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAAAA**AAA**TTTTTTTT**GA CG**AA**CG**TT**CGGCCTCAGTGAGCGAGCGAGCGCGC	-58.9	2
258		GCGCGCTCGCTCGCTCACTGAGGCC**AAAAAAAA**AAA**TTTTTTTT**GA **A**GC**AATT**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-59.3	2
259		GCGCGCTCGCTCGCTCACTGAGGCCGGGC**AGAA**AAA**TTCT**GCCCGA C**A**CCCGGG**T**GGCCTCAGTGAGCGAGCGAGCGCGC	-70.4	1
260		GCGCGCTCGCTCGCTCACTGAGGCCGGGC**AGAA**AAA**TTCT**GCCCGA CGCC**AT**GGCGGCCTCAGTGAGCGAGCGAGCGCGC	-71.9	1
261		GCGCGCTCGCTCGCTCACTGAGGCCGGGC**AGAA**AAA**TTCT**GCCCGA CG**A**C**AT**G**T**CGGCCTCAGTGAGCGAGCGAGCGCGC	-70	1
262		GCGCGCTCGCTCGCTCACTGAGGCCGGGC**AGAA**AAA**TTCT**GCCCGA CG**AA**CG**TT**CGGCCTCAGTGAGCGAGCGAGCGCGC	-70.4	1
263		GCGCGCTCGCTCGCTCACTGAGGCCGGGC**AGAA**AAA**TTCT**GCCCGA **A**GC**AATT**GC**T**GCCTCAGTGAGCGAGCGAGCGCGC	-70.8	1

Table 4: ITR Sequences with Modifications in Loop Arm and Spacer

Table 4				
SEQ ID	**Modified Region**	**Sequence**	**ΔG**	**No. Strut.**
264	Loop Arm & Spacer	GCGCGCTCGCTCGCTCACT**A**AGGCC**A**GGCGACCAAAGGTCGCC**T**GA CGCCCGGGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-71.4	1
265		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGCG**G**CCAAAGG**C**CGCCCGA CGCCCGGGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-75	2
266		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGCGAC**A**AAA**T**GTCGCCCGA CGCCCGGGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-71.2	1
267		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAA**GACCAAAGGTC**TTTT**GAC GCCCGGGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-65	2

268		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAAAAAA**AAA**TTTTTTTT**GAC GCCCGGGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-58.1	1
269		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGC**AGAA**AAA**TTCT**GCCCGA CGCCCGGGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-69.6	1
270		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**A**GGCGACCAAAGGTCGCC**T**GA CGCCCGGGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-72.3	2
271		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCGA CGCCCGGGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-75.9	3
272		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCGA CGCCCGGGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-72.1	2
273		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA CGCCCGGGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-65.9	3
274		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA CGCCCGGGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-59	2
275		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCGA CGCCCGGGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-70.5	2
276		GCGCGCTCGCTCGCTC**GAGAGAA**CC**A**GGCGACCAAAGGTCGCC**T**GA CGCCCGGGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-70.9	1
277		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCG**G**CCAAAGG**C**CGCCCG ACGCCCGGGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-74.5	1
278		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGAC**A**AAA**T**GTCGCCCGA CGCCCGGGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-70.7	1
279		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAA**GACCAAAGGTC**TTTT**GA CGCCCGGGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-64.5	2
280		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAAAAAA**AAA**TTTTTTTT**GA CGCCCGGGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-57.6	1
281		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGC**AGAA**AAA**TTCT**GCCCGA CGCCCGGGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-69.1	1
282		GCGCGCTCGCTCGCTC**G**C**G**GG**G**GCC**A**GGCGACCAAAGGTCGCC**T**GA CGCCCGGGCGGCC**CCCG**C**G**AGCGAGCGAGCGCGC	-78.8	2
283		GCGCGCTCGCTCGCTC**G**C**G**GG**G**GCCGGGCG**G**CCAAAGG**C**CGCCCG ACGCCCGGGCGGCC**CCCG**C**G**AGCGAGCGAGCGCGC	-82.4	3
284		GCGCGCTCGCTCGCTC**G**C**G**GG**G**GCCGGGCGAC**A**AAA**T**GTCGCCCGA CGCCCGGGCGGCC**CCCG**C**G**AGCGAGCGAGCGCGC	-78.6	2
285		GCGCGCTCGCTCGCTC**G**C**G**GG**G**GCC**AAAA**GACCAAAGGTC**TTTT**GA CGCCCGGGCGGCC**CCCG**C**G**AGCGAGCGAGCGCGC	-72.4	3
286		GCGCGCTCGCTCGCTC**G**C**G**GG**G**GCC**AAAAAAAA**AAA**TTTTTTTT**GA CGCCCGGGCGGCC**CCCG**C**G**AGCGAGCGAGCGCGC	-65.5	1
287		GCGCGCTCGCTCGCTC**G**C**G**GG**G**GCCGGGC**AGAA**AAA**TTCT**GCCCGA CGCCCGGGCGGCC**CCCG**C**G**AGCGAGCGAGCGCGC	-77	2
288		GCGCGCTCGCTCGCTCA**A**T**A**A**A**CC**A**GGCGACCAAAGGTCGCC**T**GA CGCCCGGGCGG**TT**T**T**AT**T**GAGCGAGCGAGCGCGC	-64.3	1
289		GCGCGCTCGCTCGCTCA**A**T**A**A**A**CCGGGCG**G**CCAAAGG**C**CGCCCGA CGCCCGGGCGG**TT**T**T**AT**T**GAGCGAGCGAGCGCGC	-67.9	1
290		GCGCGCTCGCTCGCTCA**A**T**A**A**A**CCGGGCGAC**A**AAA**T**GTCGCCCGA CGCCCGGGCGG**TT**T**T**AT**T**GAGCGAGCGAGCGCGC	-64.1	1
291		GCGCGCTCGCTCGCTCA**A**T**A**A**A**CC**AAAA**GACCAAAGGTC**TTTT**GAC GCCCGGGCGG**TT**T**T**AT**T**GAGCGAGCGAGCGCGC	-57.9	2
292		GCGCGCTCGCTCGCTCA**A**T**A**A**A**CC**AAAAAAAA**AAA**TTTTTTTT**GAC GCCCGGGCGG**TT**T**T**AT**T**GAGCGAGCGAGCGCGC	-51	1

293		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGC**AGAA**AAA**TTCT**GCCCGA CGCCCGGGCGG**TTTT**A**T**TGAGCGAGCGAGCGCGC	-62.5	1

Table 5: ITR Sequences with Modifications in Truncated Arm and Spacer

Table 5				
SEQ ID	**Modified Region**	**Sequence**	**ΔG**	**No. Strut.**
294	Truncated Arm & Spacer	GCGCGCTCGCTCGCTCACT**A**AGGCCGGGCGACCAAAGGTCGCCCGA C**A**CCCGGG**T**GGCCT**T**AGTGAGCGAGCGAGCGCGC	-71.4	1
295		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGCGACCAAAGGTCGCCCGA CGCC**AT**GGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-72.9	1
296		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGCGACCAAAGGTCGCCCGA CG**A**C**AT**G**T**CGGCCT**T**AGTGAGCGAGCGAGCGCGC	-71	1
297		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGCGACCAAAGGTCGCCCGA CG**AA**CG**TT**CGGCCT**T**AGTGAGCGAGCGAGCGCGC	-71.4	1
298		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGCGACCAAAGGTCGCCCGA **A**GC**AATT**GC**T**GCCT**T**AGTGAGCGAGCGAGCGCGC	-71.8	1
299		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGACCAAAGGTCGCCCG AC**A**CCCGGG**T**GGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-72.3	2
300		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGACCAAAGGTCGCCCG ACGCC**AT**GGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-73.8	1
301		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGACCAAAGGTCGCCCG ACG**A**C**AT**G**T**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-71.9	1
302		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGACCAAAGGTCGCCCG ACG**AA**CG**TT**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-72.3	1
303		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGACCAAAGGTCGCCCG A**A**GC**AATT**GC**T**GCC**T**CA**TC**GAGCGAGCGAGCGCGC	-72.7	1
304		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGACCAAAGGTCGCCCG AC**A**CCCGGG**T**GG**TTCTCTC**GAGCGAGCGAGCGCGC	-70.9	1
305		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGACCAAAGGTCGCCCG ACGCC**AT**GGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-72.4	1
306		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGACCAAAGGTCGCCCG ACG**A**C**AT**G**T**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-70.5	1
307		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGACCAAAGGTCGCCCG ACG**AA**CG**TT**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-70.9	1
308		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGACCAAAGGTCGCCCG A**A**GC**AATT**GC**T**G**TTCTCTC**GAGCGAGCGAGCGCGC	-71.3	1
309		GCGCGCTCGCTCGCTC**GC**G**G**GGCCGGGCGACCAAAGGTCGCCCG AC**A**CCCGGG**T**GGCC**CCC**G**C**GAGCGAGCGAGCGCGC	-78.8	1
310		GCGCGCTCGCTCGCTC**GC**G**G**GGCCGGGCGACCAAAGGTCGCCCG ACGCC**AT**GGCGGCC**CCC**G**C**GAGCGAGCGAGCGCGC	-80.3	1
311		GCGCGCTCGCTCGCTC**GC**G**G**GGCCGGGCGACCAAAGGTCGCCCG ACG**A**C**AT**G**T**CGGCC**CCC**G**C**GAGCGAGCGAGCGCGC	-78.4	1
312		GCGCGCTCGCTCGCTC**GC**G**G**GGCCGGGCGACCAAAGGTCGCCCG ACG**AA**CG**TT**CGGCC**CCC**G**C**GAGCGAGCGAGCGCGC	-78.8	1
313		GCGCGCTCGCTCGCTC**GC**G**G**GGCCGGGCGACCAAAGGTCGCCCG A**A**GC**AATT**GC**T**GCC**CCC**G**C**GAGCGAGCGAGCGCGC	-79.2	1
314		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCGACCAAAGGTCGCCCGA	-64.3	1

		CACCCGGGTGGTTTTATTGAGCGAGCGAGCGCGC		
315		GCGCGCTCGCTCGCTCAATAAAACCGGGCGACCAAAGGTCGCCCGA CGCCATGGCGGTTTTATTGAGCGAGCGAGCGCGC	-65.8	1
316		GCGCGCTCGCTCGCTCAATAAAACCGGGCGACCAAAGGTCGCCCGA CGACATGTCGGTTTTATTGAGCGAGCGAGCGCGC	-63.9	1
317		GCGCGCTCGCTCGCTCAATAAAACCGGGCGACCAAAGGTCGCCCGA CGAACGTTCGGTTTTATTGAGCGAGCGAGCGCGC	-64.3	1
318		GCGCGCTCGCTCGCTCAATAAAACCGGGCGACCAAAGGTCGCCCGA AGCAATTGCTGTTTTATTGAGCGAGCGAGCGCGC	-64.7	1

Table 6: ITR Sequences with Modifications in Loop Arm, Truncated Arm and Spacer

Table 6				
SEQ ID	**Modified Region**	**Sequence**	**ΔG**	**No. Strut.**
319	Loop Arm, Truncated Arm & Spacer	GCGCGCTCGCTCGCTCACTAAGGCCAGGCGACCAAAGGTCGCCTGA CACCCGGGTGGCCTTAGTGAGCGAGCGAGCGCGC	-69.9	2
320		GCGCGCTCGCTCGCTCACTAAGGCCAGGCGACCAAAGGTCGCCTGA CGCCATGGCGGCCTTAGTGAGCGAGCGAGCGCGC	-71.4	1
321		GCGCGCTCGCTCGCTCACTAAGGCCAGGCGACCAAAGGTCGCCTGA CGACATGTCGGCCTTAGTGAGCGAGCGAGCGCGC	-69.5	1
322		GCGCGCTCGCTCGCTCACTAAGGCCAGGCGACCAAAGGTCGCCTGA CGAACGTTCGGCCTTAGTGAGCGAGCGAGCGCGC	-69.9	1
323		GCGCGCTCGCTCGCTCACTAAGGCCAGGCGACCAAAGGTCGCCTGA AGCAATTGCTGCCTTAGTGAGCGAGCGAGCGCGC	-70.3	1
324		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGGCCAAAGGCCGCCCGA CACCCGGGTGGCCTTAGTGAGCGAGCGAGCGCGC	-73.5	2
325		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGGCCAAAGGCCGCCCGA CGCCATGGCGGCCTTAGTGAGCGAGCGAGCGCGC	-75	1
326		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGGCCAAAGGCCGCCCGA CGACATGTCGGCCTTAGTGAGCGAGCGAGCGCGC	-73.1	1
327		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGGCCAAAGGCCGCCCGA CGAACGTTCGGCCTTAGTGAGCGAGCGAGCGCGC	-73.5	1
328		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGGCCAAAGGCCGCCCGA AGCAATTGCTGCCTTAGTGAGCGAGCGAGCGCGC	-73.9	1
329		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGACAAAATGTCGCCCGA CACCCGGGTGGCCTTAGTGAGCGAGCGAGCGCGC	-69.7	1
330		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGACAAAATGTCGCCCGA CGCCATGGCGGCCTTAGTGAGCGAGCGAGCGCGC	-71.2	1
331		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGACAAAATGTCGCCCGA CGACATGTCGGCCTTAGTGAGCGAGCGAGCGCGC	-69.3	2
332		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGACAAAATGTCGCCCGA CGAACGTTCGGCCTTAGTGAGCGAGCGAGCGCGC	-69.7	2
333		GCGCGCTCGCTCGCTCACTAAGGCCGGGCGACAAAATGTCGCCCGA AGCAATTGCTGCCTTAGTGAGCGAGCGAGCGCGC	-70.1	1
334		GCGCGCTCGCTCGCTCACTAAGGCCAAAAGACCAAAGGTCTTTTGA CACCCGGGTGGCCTTAGTGAGCGAGCGAGCGCGC	-63.5	2
335		GCGCGCTCGCTCGCTCACTAAGGCCAAAAGACCAAAGGTCTTTTGA CGCCATGGCGGCCTTAGTGAGCGAGCGAGCGCGC	-65	2

336		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAA**GACCAAAGGTC**TTTT**GA CG**A**C**AT**G**T**CGGCCT**T**AGTGAGCGAGCGAGCGCGC	-63.1	2
337		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAA**GACCAAAGGTC**TTTT**GA CG**AA**CG**TT**CGGCCT**T**AGTGAGCGAGCGAGCGCGC	-63.5	2
338		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAA**GACCAAAGGTC**TTTT**GA **A**GC**AATT**GC**T**GCCT**T**AGTGAGCGAGCGAGCGCGC	-63.9	1
339		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAAAAAA**AAA**TTTTTTTT**GAC **A**CCCGGG**T**GGCCT**T**AGTGAGCGAGCGAGCGCGC	-57.3	2
340		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAAAAAA**AAA**TTTTTTTT**GAC GCC**AT**GGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-58.1	1
341		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAAAAAA**AAA**TTTTTTTT**GAC G**A**C**AT**G**T**CGGCCT**T**AGTGAGCGAGCGAGCGCGC	-57.5	1
342		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAAAAAA**AAA**TTTTTTTT**GAC G**AA**CG**TT**CGGCCT**T**AGTGAGCGAGCGAGCGCGC	-56.6	2
343		GCGCGCTCGCTCGCTCACT**A**AGGCC**AAAAAAAA**AAA**TTTTTTTT**GA **A**GC**AATT**GC**T**GCCT**T**AGTGAGCGAGCGAGCGCGC	-57	2
344		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGC**AGAA**AAA**TTCT**GCCCGA C**A**CCCGGG**T**GGCCT**T**AGTGAGCGAGCGAGCGCGC	-68.1	1
345		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGC**AGAA**AAA**TTCT**GCCCGA CGCC**AT**GGCGGCCT**T**AGTGAGCGAGCGAGCGCGC	-69.6	1
346		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGC**AGAA**AAA**TTCT**GCCCGA CG**A**C**AT**G**T**CGGCCT**T**AGTGAGCGAGCGAGCGCGC	-67.7	1
347		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGC**AGAA**AAA**TTCT**GCCCGA CG**AA**CG**TT**CGGCCT**T**AGTGAGCGAGCGAGCGCGC	-68.1	1
348		GCGCGCTCGCTCGCTCACT**A**AGGCCGGGC**AGAA**AAA**TTCT**GCCCGA **A**GC**AATT**GC**T**GCCT**T**AGTGAGCGAGCGAGCGCGC	-68.5	1
349		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**A**GGCGACCAAAGGTCGCC**T**G AC**A**CCCGGG**T**GGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-70.8	3
350		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**A**GGCGACCAAAGGTCGCC**T**G ACGCC**AT**GGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-72.3	1
351		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**A**GGCGACCAAAGGTCGCC**T**G ACG**A**C**AT**G**T**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-70.4	1
352		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**A**GGCGACCAAAGGTCGCC**T**G ACG**AA**CG**TT**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-70.8	1
353		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**A**GGCGACCAAAGGTCGCC**T**G A**A**GC**AATT**GC**T**GCC**T**CA**TC**GAGCGAGCGAGCGCGC	-71.2	1
354		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG AC**A**CCCGGG**T**GGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-74.4	3
355		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG ACGCC**AT**GGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-75.9	1
356		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG ACG**A**C**AT**G**T**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-74	1
357		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG ACG**AA**CG**TT**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-74.4	1
358		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG A**A**GC**AATT**GC**T**GCC**T**CA**TC**GAGCGAGCGAGCGCGC	-74.8	1
359		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG AC**A**CCCGGG**T**GGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-70.6	2
360		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG ACGCC**AT**GGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-72.1	1

361		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG ACG**A**C**AT**G**T**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-70.2	2
362		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG ACG**AA**CG**TT**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-70.6	2
363		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG A**A**GC**AATT**GC**T**GCC**T**CA**TC**GAGCGAGCGAGCGCGC	-71	1
364		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA C**A**CCCGGG**T**GGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-64.4	3
365		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA CGCC**AT**GGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-65.9	2
366		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA CG**A**C**AT**G**T**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-64	2
367		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA CG**AA**CG**TT**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-64.4	2
368		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA **A**GC**AATT**GC**T**GCC**T**CA**TC**GAGCGAGCGAGCGCGC	-64.8	1
369		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA C**A**CCCGGG**T**GGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-58.2	2*
370		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA CGCC**AT**GGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-59	1
371		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA CG**A**C**AT**G**T**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-58.4	1
372		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA CG**AA**CG**TT**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-57.5	2
373		GCGCGCTCGCTCGCTC**GA**TG**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA **A**GC**AATT**GC**T**GCC**T**CA**TC**GAGCGAGCGAGCGCGC	-57.9	2
374		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCGA C**A**CCCGGG**T**GGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-69	2
375		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCGA CGCC**AT**GGCGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-70.5	1
376		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCGA CG**A**C**AT**G**T**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-68.6	1
377		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCGA CG**AA**CG**TT**CGGCC**T**CA**TC**GAGCGAGCGAGCGCGC	-69	1
378		GCGCGCTCGCTCGCTC**GA**TG**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCGA **A**GC**AATT**GC**T**GCC**T**CA**TC**GAGCGAGCGAGCGCGC	-69.4	1
379		GCGCGCTCGCTCGCTC**GAGAGAA**CC**A**GGCGACCAAAGGTCGCC**T**G AC**A**CCCGGG**T**GG**TTCTCTC**GAGCGAGCGAGCGCGC	-69.4	2
380		GCGCGCTCGCTCGCTC**GAGAGAA**CC**A**GGCGACCAAAGGTCGCC**T**G ACGCC**AT**GGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-70.9	1
381		GCGCGCTCGCTCGCTC**GAGAGAA**CC**A**GGCGACCAAAGGTCGCC**T**G ACG**A**C**AT**G**T**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-69	1
382		GCGCGCTCGCTCGCTC**GAGAGAA**CC**A**GGCGACCAAAGGTCGCC**T**G ACG**AA**CG**TT**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-69.4	1
383		GCGCGCTCGCTCGCTC**GAGAGAA**CC**A**GGCGACCAAAGGTCGCC**T**G A**A**GC**AATT**GC**T**G**TTCTCTC**GAGCGAGCGAGCGCGC	-69.8	1
384		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCG**G**CCAAAGG**C**CGCCCG AC**A**CCCGGG**T**GG**TTCTCTC**GAGCGAGCGAGCGCGC	-73	1
385		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCG**G**CCAAAGG**C**CGCCCG ACGCC**AT**GGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-74.5	1

386		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCG**G**CCAAAGG**C**CGCCCG ACG**A**C**AT**G**T**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-72.6	1
387		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCG**G**CCAAAGG**C**CGCCCG ACG**AA**CG**TT**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-73	1
388		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCG**G**CCAAAGG**C**CGCCCG A**A**GC**AATT**GC**T**G**TTCTCTC**GAGCGAGCGAGCGCGC	-73.4	1
389		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGAC**A**AAA**T**GTCGCCCG AC**A**CCCGGG**T**GG**TTCTCTC**GAGCGAGCGAGCGCGC	-69.2	1
390		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGAC**A**AAA**T**GTCGCCCG ACGCC**AT**GGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-70.7	1
391		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGAC**A**AAA**T**GTCGCCCG ACG**A**C**AT**G**T**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-69.8	2
392		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGAC**A**AAA**T**GTCGCCCG ACG**AA**CG**TT**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-69.2	2
393		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGCGAC**A**AAA**T**GTCGCCCG A**A**GC**AATT**GC**T**G**TTCTCTC**GAGCGAGCGAGCGCGC	-69.6	1
394		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAA**GACCAAAGGTC**TTTT**GA C**A**CCCGGG**T**GG**TTCTCTC**GAGCGAGCGAGCGCGC	-63	2
395		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAA**GACCAAAGGTC**TTTT**GA CGCC**AT**GGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-64.5	2
396		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAA**GACCAAAGGTC**TTTT**GA CG**A**C**AT**G**T**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-62.6	2
397		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAA**GACCAAAGGTC**TTTT**GA CG**AA**CG**TT**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-63	2
398		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAA**GACCAAAGGTC**TTTT**GA **A**GC**AATT**GC**T**G**TTCTCTC**GAGCGAGCGAGCGCGC	-63.4	1
399		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAAAAAA**AAA**TTTTTTTT**GA C**A**CCCGGG**T**GG**TTCTCTC**GAGCGAGCGAGCGCGC	-56.8	2
400		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAAAAAA**AAA**TTTTTTTT**GA CGCC**AT**GGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-57.6	1
401		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAAAAAA**AAA**TTTTTTTT**GA CG**A**C**AT**G**T**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-57	1
402		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAAAAAA**AAA**TTTTTTTT**GA CG**AA**CG**TT**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-56.1	2
403		GCGCGCTCGCTCGCTC**GAGAGAA**CC**AAAAAAAA**AAA**TTTTTTTT**GA **A**GC**AATT**GC**T**G**TTCTCTC**GAGCGAGCGAGCGCGC	-56.5	2
404		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGC**AGAA**AAA**TTCT**GCCCG AC**A**CCCGGG**T**GG**TTCTCTC**GAGCGAGCGAGCGCGC	-67.6	1
405		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGC**AGAA**AAA**TTCT**GCCCG ACGCC**AT**GGCGG**TTCTCTC**GAGCGAGCGAGCGCGC	-69.1	1
406		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGC**AGAA**AAA**TTCT**GCCCG ACG**A**C**AT**G**T**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-67.2	1
407		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGC**AGAA**AAA**TTCT**GCCCG ACG**AA**CG**TT**CGG**TTCTCTC**GAGCGAGCGAGCGCGC	-67.6	1
408		GCGCGCTCGCTCGCTC**GAGAGAA**CCGGGC**AGAA**AAA**TTCT**GCCCG A**A**GC**AATT**GC**T**G**TTCTCTC**GAGCGAGCGAGCGCGC	-68	1
409		GCGCGCTCGCTCGCTC**GCGGG**GGCC**A**GGCGACCAAAGGTCGCC**T**G AC**A**CCCGGG**T**GGCC**CCCG**C**G**AGCGAGCGAGCGCGC	-77.3	2
410		GCGCGCTCGCTCGCTC**GCGGG**GGCC**A**GGCGACCAAAGGTCGCC**T**G ACGCC**AT**GGCGGCC**CCCG**C**G**AGCGAGCGAGCGCGC	-78.8	1

411		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**A**GGCGACCAAAGGTCGCC**T**G ACG**A**C**AT**G**T**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-76.9	1
412		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**A**GGCGACCAAAGGTCGCC**T**G ACG**AA**CG**TT**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-77.3	1
413		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**A**GGCGACCAAAGGTCGCC**T**G A**A**GC**AATT**GC**T**GCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-77.7	1
414		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG AC**A**CCCGGG**T**GGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-80.9	2
415		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG ACGCC**AT**GGCGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-82.4	1
416		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG ACG**A**C**AT**G**T**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-80.5	1
417		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG ACG**AA**CG**TT**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-80.9	1
418		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCG**G**CCAAAGG**C**CGCCCG A**A**GC**AATT**GC**T**GCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-81.3	1
419		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG AC**A**CCCGGG**T**GGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-77.1	1
420		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG ACGCC**AT**GGCGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-78.6	1
421		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG ACG**A**C**AT**G**T**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-76.7	2
422		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG ACG**AA**CG**TT**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-77.1	2
423		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGCGAC**A**AAA**T**GTCGCCCG A**A**GC**AATT**GC**T**GCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-77.5	1
424		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA C**A**CCCGGG**T**GGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-70.9	3
425		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA CGCC**AT**GGCGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-72.4	2
426		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA CG**A**C**AT**G**T**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-70.5	2
427		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA CG**AA**CG**TT**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-70.9	2
428		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAA**GACCAAAGGTC**TTTT**GA **A**GC**AATT**GC**T**GCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-71.3	1
429		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA C**A**CCCGGG**T**GGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-64.7	2
430		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA CGCC**AT**GGCGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-65.5	1
431		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA CG**A**C**AT**G**T**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-64.9	1
432		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA CG**AA**CG**TT**CGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-64	2
433		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCC**AAAAAAAA**AAA**TTTTTTTT**GA **A**GC**AATT**GC**T**GCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-64.4	2
434		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCG AC**A**CCCGGG**T**GGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-75.5	1
435		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCG ACGCC**AT**GGCGGCC**C**C**C**G**C**GAGCGAGCGAGCGCGC	-77	1

436		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCG ACG**A**C**AT**G**T**CGGCC**CCC**G**C**G**AGCGAGCGAGCGCGC	-75.1	1
437		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCG ACG**AA**CG**TT**CGGCC**CCC**G**C**GAGCGAGCGAGCGCGC	-75.5	1
438		GCGCGCTCGCTCGCTC**G**C**G**G**G**GGCCGGGC**AGAA**AAA**TTCT**GCCCG A**A**GC**AATT**GC**T**GCC**CCC**G**C**G**AGCGAGCGAGCGCGC	-75.9	1
439		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**A**GGCGACCAAAGGTCGCC**T**GA C**A**CCCGGG**T**GG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-62.8	2
440		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**A**GGCGACCAAAGGTCGCC**T**GA CGCC**AT**GGCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-64.3	1
441		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**A**GGCGACCAAAGGTCGCC**T**GA CG**AC**A**T**GTCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-62.4	1
442		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**A**GGCGACCAAAGGTCGCC**T**GA CG**AA**CG**TT**CGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-62.8	1
443		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**A**GGCGACCAAAGGTCGCC**T**GA **A**GC**AATT**GC**T**G**TTTT**A**TT**GAGCGAGCGAGCGCGC	-63.2	1
444		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCG**G**CCAAAGG**C**CGCCCGA C**A**CCCGGG**T**GG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-66.4	1
445		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCG**G**CCAAAGG**C**CGCCCGA CGCC**AT**GGCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-67.9	1
446		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCG**G**CCAAAGG**C**CGCCCGA CG**AC**A**T**GTCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-66	1
447		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCG**G**CCAAAGG**C**CGCCCGA CG**AA**CG**TT**CGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-66.4	1
448		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCG**G**CCAAAGG**C**CGCCCGA **A**GC**AATT**GC**T**G**TTTT**A**TT**GAGCGAGCGAGCGCGC	-66.8	1
449		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCGAC**A**AAA**T**GTCGCCCGA C**A**CCCGGG**T**GG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-62.6	1
450		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCGAC**A**AAA**T**GTCGCCCGA CGCC**AT**GGCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-64.1	1
451		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCGAC**A**AAA**T**GTCGCCCGA CG**AC**A**T**GTCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-62.2	2
452		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCGAC**A**AAA**T**GTCGCCCGA CG**AA**CG**TT**CGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-62.6	2
453		GCGCGCTCGCTCGCTCA**AT**A**AA**CCGGGCGAC**A**AAA**T**GTCGCCCGA **A**GC**AATT**GC**T**G**TTTT**A**TT**GAGCGAGCGAGCGCGC	-63	1
454		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**AAAA**GACCAAAGGTC**TTTT**GA C**A**CCCGGG**T**GG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-56.4	2
455		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**AAAA**GACCAAAGGTC**TTTT**GA CGCC**AT**GGCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-57.9	2
456		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**AAAA**GACCAAAGGTC**TTTT**GA CG**AC**A**T**GTCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-56	2
457		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**AAAA**GACCAAAGGTC**TTTT**GA CG**AA**CG**TT**CGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-56.4	2
458		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**AAAA**GACCAAAGGTC**TTTT**GA **A**GC**AATT**GC**T**G**TTTT**A**TT**GAGCGAGCGAGCGCGC	-56.8	1
459		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**AAAAAAAA**AAA**TTTTTTTT**GA C**A**CCCGGG**T**GG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-50.2	2
460		GCGCGCTCGCTCGCTCA**AT**A**AA**CC**AAAAAAAA**AAA**TTTTTTTT**GA CGCC**AT**GGCGG**TTTT**A**TT**GAGCGAGCGAGCGCGC	-51	1

461		GCGCGCTCGCTCGCTCAATAAAACCAAAAAAAAAAAATTTTTTTTGA CGACATGTCGGTTTTATTGAGCGAGCGAGCGCGC	-50.4	1
462		GCGCGCTCGCTCGCTCAATAAAACCAAAAAAAAAAATTTTTTTTGA CGAACGTTCGGTTTTATTGAGCGAGCGAGCGCGC	-49.5	2
463		GCGCGCTCGCTCGCTCAATAAAACCAAAAAAAAAAATTTTTTTTGA AGCAATTGCTGTTTTATTGAGCGAGCGAGCGCGC	-49.9	1
464		GCGCGCTCGCTCGCTCAATAAAACCGGGCAGAAAAATTCTGCCCGA CACCCGGGTGGTTTTATTGAGCGAGCGAGCGCGC	-61	1
465		GCGCGCTCGCTCGCTCAATAAAACCGGGCAGAAAAATTCTGCCCGA CGCCATGGCGGTTTTATTGAGCGAGCGAGCGCGC	-62.5	1
466		GCGCGCTCGCTCGCTCAATAAAACCGGGCAGAAAAATTCTGCCCGA CGACATGTCGGTTTTATTGAGCGAGCGAGCGCGC	-60.6	1
467		GCGCGCTCGCTCGCTCAATAAAACCGGGCAGAAAAATTCTGCCCGA CGAACGTTCGGTTTTATTGAGCGAGCGAGCGCGC	-61	1
468		GCGCGCTCGCTCGCTCAATAAAACCGGGCAGAAAAATTCTGCCCGA AGCAATTGCTGTTTTATTGAGCGAGCGAGCGCGC	-61.4	1

[00236] As disclosed herein, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can be generated to include deletion, insertion, or substitution of one or more nucleotides from the wild-type ITR derived from AAV genome. The modified ITR can be generated by genetic modification during propagation in a plasmid in *Escherichia coli* or as a baculovirus genome in *Spodoptera frugiperda* cells, or other biological methods, for example in vitro using polymerase chain reaction, or chemical synthesis.

[00237] In some embodiments, a modified ITR in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can include deletion, insertion, or substitution of one or more nucleotides from the wild-type ITR of AAV2 (Left) (SEQ ID NO: 51) or the wild-type ITR of AAV2 (Right) (SEQ ID NO: 1). Specifically, one or more nucleotides are deleted, inserted, or substituted from B-C' or C-C' of the T-shaped stem-loop structure. Furthermore, the modified ITR includes no modification in the Rep-binding elements (RBE) and the terminal resolution site (*trs*) of wild-type ITR of AAV2, although the RBE'(TTT) may be or may not be present depending on the whether the template has undergone one round of replication thereby converting the AAA triplet to the complimentary RBE' – TTT.

[00238] Three types of modified ITRs are exemplified – (1) a modified ITR having a lowest energy structure comprising a single arm and a single unpaired loop ("single-arm/single-unpaired-loop structure"); (2) a modified ITR having a lowest energy structure with a single hairpin ("single-hairpin structure"); and (3) a modified ITR having a lowest energy structure with two arms, one of which is truncated ("truncated structure").

[00239] Modified ITR with a single-arm/single-unpaired-loop structure

[00240] The wild-type ITR can be modified to form a secondary structure comprising a single arm and a single unpaired loop (*i.e.*, "single-arm/single-unpaired-loop structure"). Gibbs free energy (ΔG) of

unfolding of the structure can range between -85 kcal/mol and -70 kcal/mol. Exemplary structures of the modified ITRs are provided.

[00241] Modified ITRs predicted to form the single-arm/single-unpaired-loop structure can include deletion, insertion, or substitution of one or more nucleotides from the wild-type ITR in the sequences forming B and B' arm and/or C and C' arm. Modified ITR can be generated by genetic modification or biological and/or chemical synthesis.

[00242] For example, ITR-2, Left and Right provided in FIGS. 9A-9B (SEQ ID NOS:101 and 102), are generated to have deletion of two nucleotides from C-C' arm and deletion of 16 nucleotides from B-B' arm in the wild-type ITR of AAV2. Three nucleotides remaining in the B-B' arm of the modified ITR do not make a complementary pairing. Thus, ITR-2 Left and Right have the lowest energy structure with a single C-C' arm and a single unpaired loop. Gibbs free energy of unfolding the structure is predicted to be about -72.6 kcal/mol.

[00243] ITR-3 Left and Right provided in FIGS. 10A and 10B (SEQ ID NOS: 103 and 104), are generated to include 19 nucleotide deletions in C-C' arm from the wild-type ITR of AAV2. Three nucleotides remaining in the B-B' arm of the modified ITR do not make a complementary pairing. Thus, ITR-3 Left and Right have the lowest energy structure with a single B-B' arm and a single unpaired loop. Gibbs free energy of unfolding the structure is predicted to be about -74.8 kcal/mol.

[00244] ITR-4 Left and Right provided in FIGS. 11A and 11B (SEQ ID NOS: 105 and 106), are generated to include 19 nucleotide deletions in B-B' arm from the wild-type ITR of AAV2. Three nucleotides remaining in the B-B' arm of modified ITR do not make a complementary pairing. Thus, ITR-4 Left and Right have the lowest energy structure with a single C-C' arm and a single unpaired loop. Gibbs free energy of unfolding the structure is predicted to be about -76.9 kcal/mol.

[00245] ITR-10 Left and Right provided in FIGS. 12A and 12B (SEQ ID NOS: 107 and 108), are generated to include 8 nucleotide deletions in B-B' arm from the wild-type ITR of AAV2. Nucleotides remaining in the B-B' and C-C' arms make new complementary bonds between B and C' motives (ITR-10 Left) or between C and B' motives (ITR-10 Right). Thus, ITR-10 Left and Right have the lowest energy structure with a single B-C' or C-B' arm and a single unpaired loop. Gibbs free energy of unfolding the structure is predicted to be about -83.7 kcal/mol.

[00246] ITR-17 Left and Right provided in FIGS. 13A and 13B (SEQ ID NOS: 109 and 110), are generated to include 14 nucleotide deletions in C-C' arm from the wild-type ITR of AAV2. Eight nucleotides remaining in the C-C' arm do not make complementary bonds. As a result, ITR-17 Left and Right have the lowest energy structure with a single B-B' arm and a single unpaired loop. Gibbs free energy of unfolding the structure is predicted to be about -73.3 kcal/mol.

[00247] Sequences of wild-type ITR Left or Right (top) and various modified ITRs Left or Right (bottom) predicted to form the single-arm/single-unpaired-loop structure are aligned and provided below in Table 7.

[00248] Table 7: Alignment of wt-ITR and modified ITRs (ITR-2, ITR-3, ITR-4, ITR-10 and ITR-17) with a single-arm/single-unpaired-loop structure.

TABLE 7

Modified ITR (SEQ ID NO)	Sequence alignment of wild-type ITRs; WT-L ITR (SEQ ID NO: 540) or WT-R ITR (SEQ ID NO: 17) (top sequence) v. modified ITR sequences (SEQ ID NOs: 101, 102, 103, 104, 105, 106, 107, 108, 109, 110) (bottom sequences))	ΔG (kcal/ mol)
Left ITR-2 (SEQ:101)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG :::::::::::::::::::::::::::::: :: ::::::::: ::: GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGAAA--CCCGGGCGT---GCG-------- 10 20 30 40 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC ::::::::::::::::::::::: --------CCTCAGTGAGCGAGCGAGCGCGC 50 60 70	-72.6
Right ITR-2 (SEQ:102)	10 20 30 40 50 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACG--CCCGGGCGGC :::::::::::::::::::::::: : :::::: ::::::::: GCGCGCTCGCTCGCTCACTGAGGC----------GCACGCCCGGGTTTCCCGGGCGGC 10 20 30 40 60 70 80 CTCAGTGAGCGAGCGAGCGCGC ::::::::::::::::::::: CTCAGTGAGCGAGCGAGCGCGC 50 60 70	-72.6
Left ITR-3 (SEQ:103)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG :::::::::::::::::::::::::: ::::::::::::: GCGCGCTCGCTCGCTCACTGAGGCCG------------------TCGGGCGACCTTTGG 10 20 30 40 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC 50 60 70	-74.8
Right ITR-3 (SEQ:104)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCGGCCT ::: :::: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACG--------GCCT 10 20 30 40 50 70 80 CAGTGAGCGAGCGAGCGCGC :::::::::::::::::::: CAGTGAGCGAGCGAGCGCGC 60 70	-74.8
Left ITR-4 (SEQ:105)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG :: GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGG----------- 10 20 30 40 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC ::::::::::::::::::::::: --------CCTCAGTGAGCGAGCGAGCGCGC 50 60 70	-76.9
Right ITR-4	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCGGCCT :::::::::::::::::::::::::: : : :: : :::::::::::::	-76.9

(SEQ:106)	GCGCGCTCGCTCGCTCACTGAGGCCG--------ACGCCCGGGCTTTGCCCGGGCGGCCT 10 20 30 40 50 70 80 CAGTGAGCGAGCGAGCGCGC :::::::::::::::::: CAGTGAGCGAGCGAGCGCGC 60 70	
Left ITR-10 (SEQ:107)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG ::: ::: GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGC----TTT-- 10 20 30 40 50 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::: --GCCCGGCCTCAGTGAGCGAGCGAGCGCGC 60 70 80	-83.7
Right ITR-10 (SEQ:108)	10 20 30 40 50 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAG---GTCGCCCGAC----GCCCGG :::::::::::::::::::::::::::::: :::: : :::::: :::::: GCGCGCTCGCTCGCTCACTGAGGCCGGGC----AAAGCCCGACGCCCGGGCTTTGCCCGG 10 20 30 40 50 60 70 80 GCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::: GCGGCCTCAGTGAGCGAGCGAGCGCGC 60 70 80	-83.7
Left ITR-17 (SEQ:109)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG :::::::::::::::::::::::::::: ::: :::::::::::::::: GCGCGCTCGCTCGCTCACTGAGGCCG-------AAA-------CGTCGGGCGACCTTTGG 10 20 30 40 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC 50 60 70	-73.3
Right ITR-17 (SEQ:110)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCGGCCT :: :::::: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGTTT---CGGCCT 10 20 30 40 50 70 80 CAGTGAGCGAGCGAGCGCGC :::::::::::::::::: CAGTGAGCGAGCGAGCGCGC 60 70	-73.3

[00249] Modified ITR with a single-hairpin structure

[00250] The wild-type ITR can be modified to have the lowest energy structure comprising a single-hairpin structure. Gibbs free energy (ΔG) of unfolding of the structure can range between -70 kcal/mol and -40 kcal/mol. Exemplary structures of the modified ITRs are provided in FIGS. 14A and 14B.

[00251] Modified ITRs predicted to form the single hairpin structure can include deletion, insertion, or substitution of one or more nucleotides from the wild-type ITR in the sequences forming B and B' arm

and/or C and C' arm. Modified ITR can be generated by genetic modification or biological and/or chemical synthesis.

[00252] For example, ITR-6 Left and Right provided in FIGS. 14A and 14B (SEQ ID NOS: 111 and 112), include 40 nucleotide deletions in B-B' and C-C' arms from the wild-type ITR of AAV2. Nucleotides remaining in the modified ITR are predicted to form a single hairpin structure. Gibbs free energy of unfolding the structure is about -54.4 kcal/mol.

[00253] Sequences of wild-type ITR and ITR-6 (both left and right) are aligned and provide below in Table 8.

[00254] **Table 8**: Alignment of wt-ITR and modified ITR-6 with a single-hairpin structure.

TABLE 8

Modified ITR (SEQ ID NO)	Sequence alignment of wild-type ITRs; WT-L ITR (SEQ ID NO: 540) or WT-R ITR (SEQ ID NO: 543) (top sequence)) v. modified ITR-6 (SEQ ID NO: 111; ITR-6, left) (SEQ ID NO: 544, ITR-6 right) (bottom sequence)	ΔG (kcal/ mol)
Left ITR-6 (SEQ:111)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG :::::::::::::::::::::::: :::::: GCGCGCTCGCTCGCTCACTGAGGC---------AAAGCC---------- 10 20 30 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::: ----------TCAGTGAGCGAGCGCGC 40 50	-54.4
Right ITR-6 (SEQ:544)	80 70 60 50 40 30 , GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCT :::::::::::::::::::::::: :::: ::: , GCGCGCTCGCTCGCTCACTGAGGCC----------TTTG---------CCT 10 20 20 10 , CAGTGAGCGAGCGAGCGCGC (SEQ ID NO: 543) :::::::::::::::::: , CAGTGAGCGAGCGAGCGCGC (SEQ ID NO: 544) 40 50	-54.4

[00255] Modified ITR with a truncated structure

[00256] The wild-type ITR can be modified to have the lowest energy structure comprising two arms, one of which is truncated. Their Gibbs free energy (ΔG) of unfolding ranges between -90 and -70 kcal/mol. Thus, their Gibbs free energies of unfolding are lower than the wild-type ITR of AAV2.

[00257] The modified ITRs can include deletion, insertion, or substitution of one or more nucleotides from the wild-type ITR in the sequences forming B and B' arm and/or C and C' arm. In some embodiments, a modified ITR can, for example, comprise removal of all of a particular loop, *e.g.*, A-A' loop, B-B' loop or C-C' loop, or alternatively, the removal of 1, 2, 3, 4, 5, 6, 7, 8, 9 or more base pairs forming the stem of the

loop so long as the final loop at the end of the stem is still present. Modified ITR can be generated by genetic modification or biological and/or chemical synthesis.

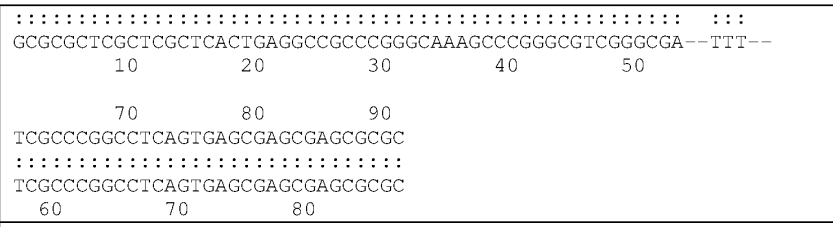
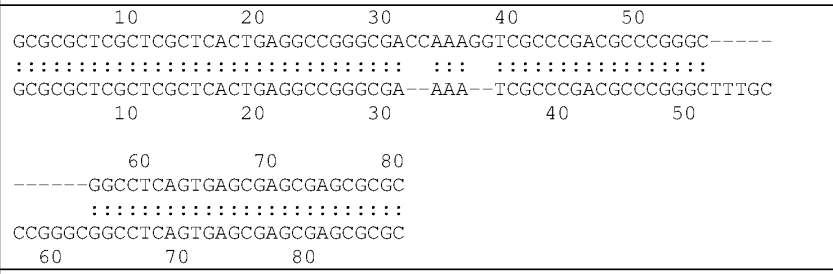
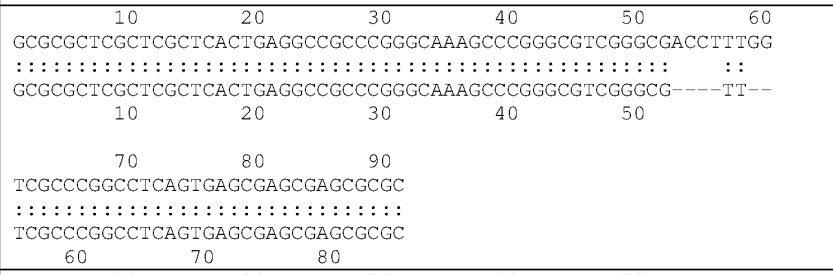
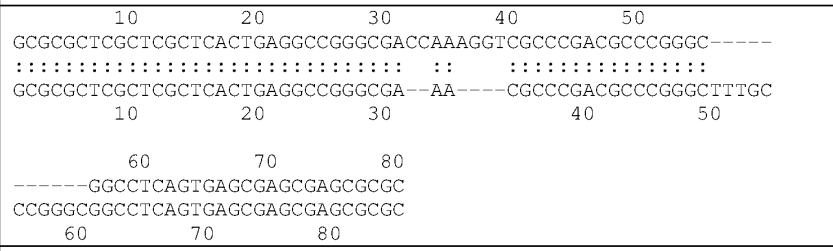
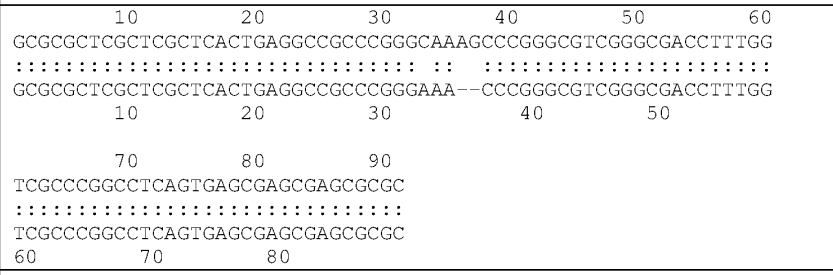
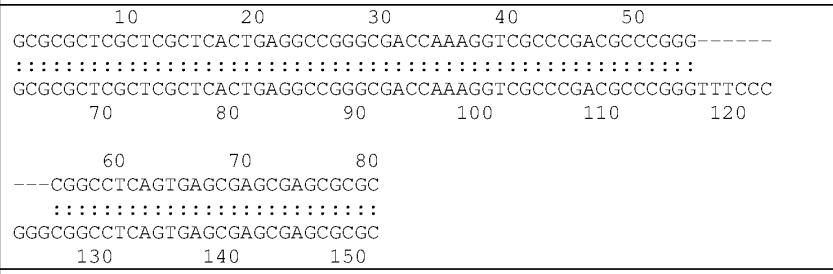
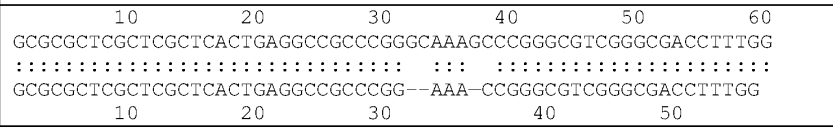
[00258] Exemplary structures of the modified ITRs with a truncated structure are provided in FIGS. 15A-15B.

[00259] Sequences of various modified ITRs predicted to form a truncated structure are aligned with a sequence of wild-type ITR and provided below in Table 9.

[00260] Table 9: Alignment of wt-ITR and modified ITRs (ITR-5, ITR-7, ITR-8, ITR-9, ITR-11, ITR-12, ITR-13, ITR-14, ITR-1 and ITR-16) with a truncated structure.

TABLE 9

Modified ITR (SEQ ID NO)	Sequence alignment: wild-type ITRs; WT-L ITR (SEQ ID NO: 540) or WT-R ITR (SEQ ID NO: 17) (top sequence) v. modified ITRs) (SEQ ID NOs: 545 and 116-134) (bottom sequences)	ΔG (kcal/mol)
Left ITR-5 (SEQ:545)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG :::::::::::::::::::::::: :::::::::::::::::::::::: GCGCGCTCGCTCGCTCACTGAGGC----------GCCCGGGCGTCGGGCGACCTTTGG 10 20 30 40 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC (SEQ ID NO: 545) 50 60 70	-73.4
Right ITR-5 (SEQ:116)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCGGCCT ::: ::: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCG-CCT 10 20 30 40 50 70 80 CAGTGAGCGAGCGAGCGCGC :::::::::::::::::::: CAGTGAGCGAGCGAGCGCGC 60 70	-73.4
Left ITR-7 (SEQ:117)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG ::: :: : GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGAC--TTTG 10 20 30 40 50 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC 60 70 80	-89.6
Right ITR-7 (SEQ:118)	10 20 30 40 50 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGC----- :::::::::::::::::::::::::::::: :: :::::::::::::::::: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACAAA--GTCGCCCGACGCCCGGGCTTTGC 10 20 30 40 50 60 70 80 ------GGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::: CCGGGCGGCCTCAGTGAGCGAGCGAGCGCGC 60 70 80	-89.6
Left ITR-8	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG	-86.9

(SEQ:119)		
Right ITR-8 (SEQ:120)		-86.9
Left ITR-9 (SEQ:121)		-85.0
Right ITR-9 (SEQ:122)		-85.0
Left ITR-11 (SEQ:123)		-89.5
Right ITR-11 (SEQ:124)		-89.5
Left ITR-12 (SEQ:125)		-86.2

	70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC 60 70 80	
Right ITR-12 (SEQ:126)	10 20 30 40 50 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGG------- :: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGTTTCCGG 10 20 30 40 50 60 60 70 80 GCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::: GCGGCCTCAGTGAGCGAGCGAGCGCGC 70 80	-86.2
Left ITR-13 (SEQ:127)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG ::::::::::::::::::::::::::::::: ::: :::::::::::::::::::: GCGCGCTCGCTCGCTCACTGAGGCCGCCCG---AAA---CGGGCGTCGGGCGACCTTTGG 10 20 30 40 50 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC 60 70 80	-82.9
Right ITR-13 (SEQ:128)	10 20 30 40 50 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCG-----GGC ::: ::: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGTTTCGGGC 10 20 30 40 50 60 60 70 80 GGCCTCAGTGAGCGAGCGAGCGCGC ::::::::::::::::::::::::: GGCCTCAGTGAGCGAGCGAGCGCGC 70 80	-82.9
Left ITR-14 (SEQ:129)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG ::::::::::::::::::::::::::::: :::: :::::::::::::::::: GCGCGCTCGCTCGCTCACTGAGGCCGCCC----AAAG----GGCGTCGGGCGACCTTTGG 10 20 30 40 50 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC 60 70 80	-80.5
Right ITR-14 (SEQ:130)	10 20 30 40 50 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCC---GGGCGG ::: :::::: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCTTTGGGCGG 10 20 30 40 50 60 60 70 80 CCTCAGTGAGCGAGCGAGCGCGC ::::::::::::::::::::::: CCTCAGTGAGCGAGCGAGCGCGC 70 80	-80.5
Left ITR-15 (SEQ:131)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG ::::::::::::::::::::::::::: :::: :::::::::::::::: GCGCGCTCGCTCGCTCACTGAGGCCGCC-----AAAG-----GCGTCGGGCGACCTTTGG 10 20 30 40 50 70 80 90	-77.2

	TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC 60 70 80	
Right ITR-15 (SEQ:132)	10 20 30 40 50 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCG-GGCGGCC ::: ::::::: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCTTTGGCGGCC 10 20 30 40 50 60 60 70 80 TCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::: TCAGTGAGCGAGCGAGCGCGC 70 80	-77.2
Left ITR-16 (SEQ:133)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG :::::::::::::::::::::::::::: ::::: :::::::::::::::: GCGCGCTCGCTCGCTCACTGAGGCCGC------AAAGC------GTCGGGCGACCTTTGG 10 20 30 40 70 80 90 TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC :::::::::::::::::::::::::::::: TCGCCCGGCCTCAGTGAGCGAGCGAGCGCGC 50 60 70	-73.9
Right ITR-16 (SEQ:134)	10 20 30 40 50 60 GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCGGCCT ::: : :::::: GCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCTTTG-CGGCCT 10 20 30 40 50 70 80 CAGTGAGCGAGCGAGCGCGC :::::::::::::::::::: CAGTGAGCGAGCGAGCGCGC 60 70	-73.9

[00261] Additional exemplary modified ITRs in each of the above classes for use in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein are provided in Tables 10A and 10B. The predicted secondary structure of the Right modified ITRs in Table 10A are shown in FIG. 26A, and the predicted secondary structure of the Left modified ITRs in Table 10B are shown in FIG. 26B.

[00262] Table 10A and Table 10B show exemplary right and left modified ITRs in a a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein.

[00263] Table 10A: Exemplary modified right ITRs. These exemplary modified right ITRs can comprise the RBE of GCGCGCTCGCTCGCTC-3′ (SEQ ID NO: 531), spacer of ACTGAGGC (SEQ ID NO: 532), the spacer complement GCCTCAGT (SEQ ID NO: 535) and RBE' (*i.e.*, complement to RBE) of GAGCGAGCGAGCGCGC (SEQ ID NO: 536).

Table 10A: Exemplary Right modified ITRs

ITR Construct	Sequence	SEQ ID NO:
ITR-18 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCGCACGCCCGGGTTTCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	469
ITR-19 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	470
ITR-20 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	471
ITR-21 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCTTTGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	472
ITR-22 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACAAAGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	473
ITR-23 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGAAAATCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	474
ITR-24 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGAAACGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	475
ITR-25 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCAAAGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	476
ITR-26 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGTTTCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	477
ITR-27 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGTTTCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	478
ITR-28 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGTTTCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	479
ITR-29 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCTTTGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	480
ITR-30 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCTTTGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	481
ITR-31 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCTTTGCGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	482

ITR-32 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCG CTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGTTTCGG CCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	483
ITR-49 Right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCG CTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGGCCTCA GTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	99
ITR-50 right	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCG CTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGG CGGCCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGG	100

[00264] TABLE 10B: Exemplary modified left ITRs in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein. These exemplary modified left ITRs can comprise the RBE of GCGCGCTCGCTCGCTC-3′ (SEQ ID NO: 531), spacer of ACTGAGGC (SEQ ID NO: 532), the spacer complement GCCTCAGT (SEQ ID NO: 535) and RBE complement (RBE') of GAGCGAGCGAGCGCGC (SEQ ID NO: 536).

Table 10B: Exemplary modified left ITRs		
ITR-33 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGG AAACCCGGGCGTGCGCCTCAGTGAGCGAGCGAGCGCGCAGAGAG GGAGTGGCCAACTCCATCACTAGGGGTTCCT	484
ITR-34 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGTCGGGC GACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGA GGGAGTGGCCAACTCCATCACTAGGGGTTCCT	485
ITR-35 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGG CAAAGCCCGGGCGTCGGCCTCAGTGAGCGAGCGAGCGCGCAGAG AGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	486
ITR-36 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCGCCCGGGC GTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGC GCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	487
ITR-37 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCAAAGCCTC AGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCA CTAGGGGTTCCT	488
ITR-38 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGG CAAAGCCCGGGCGTCGGGCGACTTTGTCGCCCGGCCTCAGTGAGC GAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGT TCCT	489
ITR-39 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGG CAAAGCCCGGGCGTCGGGCGATTTTCGCCCGGCCTCAGTGAGCGA GCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTC CT	490
ITR-40 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGG CAAAGCCCGGGCGTCGGGCGTTTCGCCCGGCCTCAGTGAGCGAGC GAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	491
ITR-41 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGG CAAAGCCCGGGCGTCGGGCTTTGCCCGGCCTCAGTGAGCGAGCGA GCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	492
ITR-42 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGG AAACCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGC GAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGT TCCT	493

ITR-43 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGAAACCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	494
ITR-44 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGAAACGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	495
ITR-45 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCAAAGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	496
ITR-46 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCAAAGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	497
ITR-47 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCAAAGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	498
ITR-48 Left	CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGAAACGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCT	499

[00265] In embodiments of the present invention, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein does not have a modified ITRs having the nucleotide sequence selected from any of the group of SEQ ID Nos: 550, 551, 552, 553, 553, 554, 555, 556, 557.

[00266] To the extent a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein has a modified ITR that has one of the modifications in the B, B', C or C' region as described in SEQ ID NO: 550-557 as defined in any one or more of the claims of this application, or within any invention to be defined in amended claims that may in the future be filed in this application or in any patent derived therefrom, and to the extent that the laws of any relevant country or countries to which that or those claims apply, we hereby reserve the right to disclaim the said disclosure from the claims of the present application or any patent derived therefrom to the extent necessary to prevent invalidation of the present application or any patent derived therefrom.

[00267] For example, and without limitation, we reserve the right to disclaim any one of the following subject-matters from any claim of the present application, now or as amended in the future, or any patent derived therefrom:

[00268] A. a modified ITR selected from any of the group consisting of: SEQ ID NOS: 2, 52, 63 64, 113, 114, 550, 551; 552, 553, 553, 554, 555, 556, 557 used in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein, without a regulatory switch

[00269] B. the above-specified modified ITRs in A., in a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein, without a regulatory sequence

and where the heterologous nucleic acid encodes ABCA4, USA2A var1, VEGFR, CEP290, BDD Factor VIII (FVIII), Factor VIII, vWF_His, vWF, lecithin cholesterol acetyl transferase, PAH, G6PC, or CFTR

VI. Regulatory elements.

[00270] A composition useful in the methods to produce a DNA vector, *e.g.*, ceDNA vector as described herein or AAV vector, comprises a nucleic acid sequence encoding a single modified Rep protein can further comprise a regulatory element, *e.g.*, a cis-regulatory element as described herein upstream to, or operatively linked to the nucleic acid encoding a single modified Rep protein. For example, a nucleotide sequence encoding a modified Rep protein, *e.g.*, encoding a modified Rep 78 protein, but not comprising a functional initiation codon for encoding the Rep 52 protein, or splice sites for exon skipping for production of Rep 68 or Rep40, is operatively linked to a regulatory element, *e.g.*, a cis-regulatory element.

[00271] In one embodiment, a nucleotide sequence encoding a single Rep protein useful in the compositions and methods as disclosed herein comprises an expression control sequence, *e.g.*, promoter, cis-regulatory elements, or regulatory switch as described herein, located upstream of the initiation codon of the nucleotide sequence encoding the parvoviral Rep78 protein, where the nucleic acid sequence does not have a functional initiation codon for Rep52 and/or splice sites for exon skipping for production of Rep 68 or Rep40. In one embodiment, a nucleotide sequence encoding a single Rep protein useful in the compositions and methods as disclosed herein comprises an expression control sequence upstream of the initiation codon of the nucleotide sequence encoding the parvoviral Rep 78 protein, where the nucleic acid sequence does not have a functional spice sites for encoding Rep68.

[00272] Similarly, a ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein can be produced from expression constructs that further comprise a specific combination of cis-regulatory elements. The cis-regulatory elements include, but are not limited to, a promoter, a riboswitch, an insulator, a mir-regulatable element, a post-transcriptional regulatory element, a tissue- and cell type-specific promoter and an enhancer. In some embodiments the ITR can act as the promoter for the transgene. In some embodiments, the ceDNA vector comprises additional components to regulate expression of the transgene, for example, regulatory switches as described herein, to regulate the expression of the transgene, or a kill switch, which can kill a cell comprising the ceDNA vector.

[00273] A ceDNA vector produced according to the methods and compositions using a single Rep protein as disclosed herein, can be produced from expression constructs that further comprise a specific combination of cis-regulatory elements such as WHP posttranscriptional regulatory element (WPRE) (*e.g.*, SEQ ID NO: 8) and BGH polyA (SEQ ID NO: 9). Suitable expression cassettes for use in expression constructs are not limited by the packaging constraint imposed by the viral capsid. Expression cassettes of the present invention include a promoter, which can influence overall expression levels as well as cell-specificity. For transgene expression, they can include a highly active virus-derived immediate early promoter. Expression cassettes can contain tissue-specific eukaryotic promoters to limit transgene

expression to specific cell types and reduce toxic effects and immune responses resulting from unregulated, ectopic expression.

[00274] In preferred embodiments, promoters or regulatory elements for use in expressing a modified single Rep protein, or in an expression cassette of a ceDNA vector produced by the methods as disclosed herein can contain a synthetic regulatory element, such as a CAG promoter (SEQ ID NO: 3). The CAG promoter comprises (i) the cytomegalovirus (CMV) early enhancer element, (ii) the promoter, the first exon and the first intron of chicken beta-actin gene, and (iii) the splice acceptor of the rabbit beta-globin gene. Alternatively, promoters or regulatory elements for use in expressing a modified single Rep protein, or in an expression cassette of a ceDNA vector produced by the methods as disclosed herein can contain an Alpha-1-antitrypsin (AAT) promoter (SEQ ID NO: 4 or SEQ ID NO: 74), a liver specific (LP1) promoter (SEQ ID NO: 5 or SEQ ID NO: 16), or a Human elongation factor-1 alpha (EF1a) promoter (*e.g.*, SEQ ID NO: 6 or SEQ ID NO: 15). In some embodiments, promoters or regulatory elements for use in expressing a modified single Rep protein, or in an expression cassette of a ceDNA vector produced by the methods as disclosed herein is selected from one or more of the constitutive promoters, for example, a retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), or a cytomegalovirus (CMV) immediate early promoter (optionally with the CMV enhancer, *e.g.*, SEQ ID NO: 22). Alternatively, an inducible promoter, a native promoter for a transgene, a tissue-specific promoter, or various promoters known in the art can be operatively linked to the nucleic acid encoding a modified single Rep protein, or in an expression cassette of a ceDNA vector produced by the methods as disclosed herein.

[00275] Suitable promoters, including those described above, can be derived from viruses and can therefore be referred to as viral promoters, or they can be derived from any organism, including prokaryotic or eukaryotic organisms. Suitable promoters can be used to drive expression by any RNA polymerase (*e.g.,* pol I, pol II, pol III). Exemplary promoters that can be operatively linked to the nucleic acid encoding a modified single Rep protein, or in an expression cassette of a ceDNA vector produced by the methods as disclosed herein, include, but are not limited to the SV40 early promoter, mouse mammary tumor virus long terminal repeat (LTR) promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), a rous sarcoma virus (RSV) promoter, a human U6 small nuclear promoter (U6, *e.g.,* SEQ ID NO: 18) (Miyagishi *et al.*, *Nature Biotechnology* 20, 497-500 (2002)), an enhanced U6 promoter (*e.g.,* Xia *et al.*, *Nucleic Acids Res*. 2003 Sep. 1; 31(17)), a human H1 promoter (H1) (*e.g.,* SEQ ID NO: 19), a CAG promoter, a human alpha 1-antitrypsin (HAAT) promoter (*e.g.*, SEQ ID NO: 21), and the like. In embodiments, these promoters are altered at their downstream intron containing end to include one or more nuclease cleavage sites. In embodiments, the DNA containing the nuclease cleavage site(s) is foreign to the promoter DNA.

[00276] A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance expression and/or to alter the spatial expression and/or temporal expression of same. A promoter

may also comprise distal enhancer or repressor elements, which may be located as much as several thousand base pairs from the start site of transcription. A promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively, or differentially with respect to the cell, tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents. Representative examples of promoters that can be operatively linked to the nucleic acid encoding a modified single Rep protein, or in an expression cassette of a ceDNA vector produced by the methods as disclosed herein, include, but are not limited to, the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter and the CMV IE promoter, as well as the promoters listed below. Such promoters and/or enhancers can be used for expression of any gene of interest, *e.g.,* the gene editing molecules, donor sequence, therapeutic proteins etc.). For example, the vector may comprise a promoter that is operably linked to the nucleic acid sequence encoding a therapeutic protein. The promoter operably linked to the therapeutic protein coding sequence may be a promoter from simian virus 40 (SV40), a mouse mammary tumor virus (MMTV) promoter, a human immunodeficiency virus (HIV) promoter such as the bovine immunodeficiency virus (BIV) long terminal repeat (LTR) promoter, a Moloney virus promoter, an avian leukosis virus (ALV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter, Epstein Barr virus (EBV) promoter, or a Rous sarcoma virus (RSV) promoter. The promoter may also be a promoter from a human gene such as human ubiquitin C (hUbC), human actin, human myosin, human hemoglobin, human muscle creatine, or human metallothionein. The promoter may also be a tissue specific promoter, such as a liver specific promoter, such as human alpha 1-antitrypsin (HAAT), natural or synthetic. In one embodiment, delivery to the liver can be achieved using endogenous ApoE specific targeting of the composition comprising a ceDNA vector to hepatocytes via the low density lipoprotein (LDL) receptor present on the surface of the hepatocyte.

[00277] In one embodiment, the promoter used is the native promoter of the gene encoding the therapeutic protein. The promoters and other regulatory sequences for the respective genes encoding the therapeutic proteins are known and have been characterized. The promoter region used may further include one or more additional regulatory sequences (*e.g.*, native), *e.g.*, enhancers, (*e.g.* SEQ ID NO: 22 and SEQ ID NO: 23).

[00278] Non-limiting examples of suitable promoters for use in expressing a modified single Rep protein, or a ceDNA vector produced by the methods as disclosed herein, include the CAG promoter of, for example (SEQ ID NO: 3), the HAAT promoter (SEQ ID NO: 21), the human EF1-α promoter (SEQ ID NO: 6) or a fragment of the EF1a promoter (SEQ ID NO: 15), IE2 promoter (*e.g.*, SEQ ID NO: 20) and the rat EF1-α promoter (SEQ ID NO: 24).

[00279] *Polyadenylation Sequences:* In some embodiments, a sequence encoding a polyadenylation sequence can be operatively linked to the nucleic acid encoding a modified single Rep protein, or in a ceDNA vector produced by the methods as disclosed herein in order to stabilize the mRNA expressed, and/or to aid in nuclear export and translation. In one embodiment, a construct comprising a nucleic acid encoding a modified single Rep protein, or a ceDNA vector produced by the methods as disclosed herein does not include a polyadenylation sequence. In alternative embodiments, a construct comprising a nucleic acid encoding a modified single Rep protein, or a ceDNA vector produced by the methods as disclosed herein includes at least 1, at least 2, at least 3, at least 4, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, least 45, at least 50 or more adenine dinucleotides. In some embodiments, the polyadenylation sequence comprises about 43 nucleotides, about 40-50 nucleotides, about 40-55 nucleotides, about 45-50 nucleotides, about 35-50 nucleotides, or any range there between.

[00280] A construct comprising a nucleic acid encoding a modified single Rep protein, or a ceDNA vector produced by the methods as disclosed herein can include a poly-adenylation sequence known in the art or a variation thereof, such as a naturally occurring sequence isolated from bovine BGHpA (*e.g.*, SEQ ID NO: 74) or a virus SV40pA (*e.g.*, SEQ ID NO: 10), or a synthetic sequence (*e.g.*, SEQ ID NO: 27). Some expression cassettes can also include SV40 late polyA signal upstream enhancer (USE) sequence. In some embodiments, the, USE can be used in combination with SV40pA or heterologous poly-A signal.

[00281] The expression cassettes can also include a post-transcriptional element to increase the expression of a transgene. In some embodiments, Woodchuck Hepatitis Virus (WHP) posttranscriptional regulatory element (WPRE) (*e.g.*, SEQ ID NO: 8) is used to increase the expression of a transgene. Other posttranscriptional processing elements such as the post-transcriptional element from the thymidine kinase gene of herpes simplex virus, or hepatitis B virus (HBV) can be used. Secretory sequences can be linked to the transgenes, *e.g.*, VH-02 and VK-A26 sequences, *e.g.*, SEQ ID NO: 25 and SEQ ID NO: 26.

VI. Regulatory Switches

[00282] A molecular regulatory switch is one which generates a measurable change in state in response to a signal. Such regulatory switches can be usefully combined with a construct comprising a nucleic acid encoding a modified single Rep protein, or a ceDNA vector produced by the methods as disclosed herein to control the output of the ceDNA vector. In some embodiments, a construct comprising a nucleic acid encoding a modified single Rep protein, or a ceDNA vector produced by the methods as disclosed herein comprises a regulatory switch that serves to fine tune the expression of the single Rep protein or the transgene in the ceDNA vector. For example, it can serve as a biocontainment function of the ceDNA vector. In some embodiments, the switch is an "ON/OFF" switch that is designed to start or stop (*i.e.*, shut down) expression of the gene of interest in the ceDNA in a controllable and regulatable fashion. In some embodiments, the switch can include a "kill switch" that can instruct the cell comprising the ceDNA vector to undergo cell programmed death once the switch is activated.

[00283] *A. Binary Regulatory Switches*

[00284] In some embodiments, the ceDNA vector comprises a regulatory switch that can serve to controllably modulate expression of the transgene. In such an embodiment, the expression cassette located between the ITRs of the ceDNA vector may additionally comprise a regulatory region, *e.g.*, a promoter, cis-element, repressor, enhancer etc., that is operatively linked to the gene of interest, where the regulatory region is regulated by one or more cofactors or exogenous agents. Accordingly, in one embodiment, only when the one or more cofactor(s) or exogenous agents are present in the cell will transcription and expression of the gene of interest from the ceDNA vector occur. In another embodiment, one or more cofactor(s) or exogenous agents may be used to de-repress the transcription and expression of the gene of interest.

[00285] Any nucleic acid regulatory regions known by a person of ordinary skill in the art can be employed in a ceDNA vector designed to include a regulatory switch. By way of example only, regulatory regions can be modulated by small molecule switches or inducible or repressible promoters. Nonlimiting examples of inducible promoters are hormone-inducible or metal-inducible promoters. Other exemplary inducible promoters/enhancer elements include, but are not limited to, an RU486-inducible promoter, an ecdysone-inducible promoter, a rapamycin-inducible promoter, and a metallothionein promoter. Classic tetracycline-based or other antibiotic-based switches are encompassed for use, including those disclosed in (Fussenegger *et al.*, Nature Biotechnol. 18: 1203-1208 (2000)).

[00286] *B. Small molecule Regulatory Switches*

[00287] A variety of art-known small-molecule based regulatory switches are known in the art and can be combined with the ceDNA vectors disclosed herein to form a regulatory-switch controlled ceDNA vector. In some embodiments, the regulatory switch can be selected from any one or a combination of: an orthogonal ligand/nuclear receptor pair, for example retinoid receptor variant/LG335 and GRQCIMFI, along with an artificial promoter controlling expression of the operatively linked transgene, such as that as disclosed in Taylor, *et al.* BMC Biotechnology 10 (2010): 15; engineered steroid receptors, *e.g.*, modified progesterone receptor with a C-terminal truncation that cannot bind progesterone but binds RU486 (mifepristone) (US Patent No. 5,364,791); an ecdysone receptor from *Drosophila* and their ecdysteroid ligands (Saez, *et al.*, PNAS, 97(26)(2000), 14512–14517; or a switch controlled by the antibiotic trimethoprim (TMP), as disclosed in Sando R 3rd; Nat Methods. 2013, 10(11):1085-8.

[00288] Other small molecule based regulatory switches known by an ordinarily skilled artisan are also envisioned for use to control transgene expression of the ceDNA and include, but are not limited to, those disclosed in Buskirk *et al.*, Cell; Chem and Biol., 2005; 12(2); 151-161; an abscisic acid sensitive ON-switch; such as that disclosed in Liang, F.-S., *et al.*, (2011) Science Signaling, 4(164);exogenous L-arginine sensitive ON-switches such as those disclosed in Hartenbach, *et al. Nucleic Acids Research*, 35(20), 2007, synthetic bile-acid sensitive ON-switches such as those disclosed in Rössger *et al.*, Metab Eng. 2014,21: 81–90;biotin sensitive ON-switches such as those disclosed in Weber *et al.*, Metab. Eng. 2009 Mar; 11(2): 117–124; dual input food additive benzoate/vanillin sensitive regulatory switches such as those disclosed in

Xie *et al.*, Nucleic Acids Research, 2014; 42(14);e116; 4-hydroxytamoxifen sensitive switches such as those disclosed in Giuseppe *et al.*, Molecular Therapy , 6(5), 653 – 663; and flavonoid (phloretin) sensitive regulatory switches such as those disclosed in Gitzinger *et al.*, Proc. Natl. Acad. Sci. U S A. 2009 Jun 30; 106(26): 10638–10643.

[00289] In some embodiments, the regulatory switch to control the transgene or expressed by the ceDNA vector is a pro-drug activation switch, such as that disclosed in US patents 8,771,679, and 6,339,070.

[00290] Exemplary regulatory switches for use in the ceDNA vectors include, but are not limited to those in Table 11.

[00291] *C. "Passcode" Regulatory Switches*

[00292] In some embodiments the regulatory switch can be a "passcode switch" or "passcode circuit". Passcode switches allow fine tuning of the control of the expression of the transgene from the ceDNA vector when specific conditions occur – that is, a combination of conditions need to be present for transgene expression and/or repression to occur. For example, for expression of a transgene to occur at least conditions A and B must occur. A passcode regulatory switch can be any number of conditions, *e.g.*, at least 2, or at least 3, or at least 4, or at least 5, or at least 6 or at least 7 or more conditions to be present for transgene expression to occur. In some embodiments, at least 2 conditions (*e.g.*, A, B conditions) need to occur, and in some embodiments, at least 3 conditions need to occur (*e.g.*, A, B and C, or A, B and D). By way of an example only, for gene expression from a ceDNA to occur that has a passcode "ABC" regulatory switch, conditions A, B and C must be present. Conditions A, B and C could be as follows; condition A is the presence of a condition or disease, condition B is a hormonal response, and condition C is a response to the transgene expression. As an exemplary example only, if the transgene is insulin, Condition A occurs if the subject has diabetes, Condition B is if the sugar level in the blood is high and Condition C is the level of endogenous insulin not being expressed at required amounts. Once the sugar level declines or the desired level of insulin is reached, the transgene (*e.g.* insulin), turns off again until the 3 conditions occur, turning it back on. In another exemplary example, if the transgene is EPO, Condition A is the presence of Chronic Kidney Disease (CKD), Condition B occurs if the subject has hypoxic conditions in the kidney, Condition C is that Erythropoietin-producing cells (EPC) recruitment in the kidney is impaired; or alternatively, HIF-2 activation is impaired. Once the oxygen levels increase or the desired level of EPO is reached, the transgene (*e.g.*, EPO) turns off again until 3 conditions occur, turning it back on.

[00293] Passcode regulatory switches are useful to fine tune the expression of the transgene from the ceDNA vector. For example, the passcode regulatory switch can be modular in that it comprises multiple switches, *e.g.*, a tissue specific, inducible promoter that is turned on only in the presence of a certain level of a metabolite. In such an embodiment, for transgene expression from the ceDNA vector to occur, the inducible agent must be present (condition A), in the desired cell type (condition B) and the metabolite is at, or above or below a certain threshold (Condition C). In alternative embodiments, the passcode regulatory switch can be designed such that the transgene expression is on when conditions A and B are present, but

will turn off when condition C is present. Such an embodiment is useful when Condition C occurs as a direct result of the expressed transgene – that is Condition C serves as a positive feedback to loop to turn off transgene expression from the ceDNA vector when the transgene has had a sufficient amount of the desired therapeutic effect.

[00294] In some embodiments, a passcode regulatory switch encompassed for use in the ceDNA vector is disclosed in WO2017/059245, incorporated by reference in its entirety herein, which describes a switch referred to as a "Passcode switch" or a "Passcode circuit" or "Passcode kill switch" which is a synthetic biological circuit that uses hybrid transcription factors (TFs) to construct complex environmental requirements for cell survival. The Passcode regulatory switches described in WO2017/059245 are particularly useful for use in the ceDNA vectors, as they are modular and customizable, both in terms of the environmental conditions that control circuit activation and in the output modules that control cell fate. In addition, the Passcode circuit has particular utility to be used in ceDNA vectors, since without the appropriate "passcode" molecules it will allow transgene expression only in the presence of the required predetermined conditions. If something goes wrong with a cell or no further transgene expression is desired for any reason, then the related kill switch (*i.e.* deadman switch) can be triggered.

[00295] In some embodiments, a passcode regulatory switch or "Passcode circuit" encompassed for use in the ceDNA vector comprises hybrid transcription factors (TFs) to expand the range and complexity of environmental signals used to define biocontainment conditions. As opposed to the deadman switch which triggers cell death on in the presence of a predetermined condition, the "passcode circuit" allows cell survival or transgene expression in the presence of a particular "passcode", and can be easily reprogrammed to allow transgene expression and/or cell survival only when the predetermined environmental condition or passcode is present.

[00296] In one aspect, a "passcode" system that restricts cell growth to the presence of a predetermined set of at least two selected agents, includes one or more nucleic acid constructs encoding expression modules comprising: i) a toxin expression module that encodes a toxin that is toxic to a host cell, wherein sequence encoding the toxin is operably linked to a promoter P1 that is repressed by the binding of a first hybrid repressor protein hRP1; ii) a first hybrid repressor protein expression module that encodes the first hybrid repressor protein hRP1, wherein expression of hRP1 is controlled by an AND gate formed by two hybrid transcription factors hTF1 and hTF2, the binding or activity of which is responsive to agents A1 and A2, respectively, such that both agents A1 and A2 are required for expression of hRP1, wherein in the absence of either A1 or A2, hRP1 expression is insufficient to repress toxin promoter module P1 and toxin production, such that the host cell is killed. In this system, hybrid factors hTF1, hTF2 and hRP1 each comprise an environmental sensing module from one transcription factor and a DNA recognition module from a different transcription factor that renders the binding of the respective passcode regulatory switch sensitive to the presence of an environmental agent, A1, or A2, that is different from that which the respective subunits would typically bind in nature.

[00297] Accordingly, a ceDNA vector can comprise a 'Passcode regulatory circuit" that requires the presence and/or absence of specific molecules to activate the output module. In some embodiments, where genes that encode for cellular toxins are placed in the output module, this passcode regulatory circuit can not only be used to regulate transgene expression, but also can be used to create a kill switch mechanism in which the circuit kills the cell if the cell behaves in an undesired fashion (*e.g.*, it leaves the specific environment defined by the sensor domains, or differentiates into a different cell type). In one nonlimiting example, the modularity of the hybrid transcription factors, the circuit architecture, and the output module allows the circuit to be reconfigured to sense other environmental signals, to react to the environmental signals in other ways, and to control other functions in the cell in addition to induced cell death, as is understood in the art.

[00298] Any and all combinations of regulatory switches disclosed herein, *e.g.*, small molecule switches, nucleic acid-based switches, small molecule-nucleic acid hybrid switches, post-transcriptional transgene regulation switches, post-translational regulation, radiation-controlled switches, hypoxia-mediated switches and other regulatory switches known by persons of ordinary skill in the art as disclosed herein can be used in a passcode regulatory switch as disclosed herein. Regulatory switches encompassed for use are also discussed in the review article Kis *et al.*, J R Soc Interface. 12: 20141000 (2015), and summarized in Table 1 of Kis. In some embodiments, a regulatory switch for use in a passcode system can be selected from any or a combination of the switches in Table 11.

[00299] *D. Nucleic acid-based regulatory switches to control transgene expression*

[00300] In some embodiments, the regulatory switch to control the transgene expressed by the ceDNA is based on a nucleic-acid based control mechanism. Exemplary nucleic acid control mechanisms are known in the art and are envisioned for use. For example, such mechanisms include riboswitches, such as those disclosed in, *e.g.*, US2009/0305253, US2008/0269258, US2017/0204477, WO2018026762A1, US patent 9,222,093 and EP application EP288071, all of which are incorporated by reference in their entireties herein, and also disclosed in the review by Villa JK *et al.*, Microbiol Spectr. 2018 May;6(3), incorporated by reference in its entirety herein. Also included are metabolite-responsive transcription biosensors, such as those disclosed in WO2018/075486 and WO2017/147585, incorporated by reference in their entireties herein. Other art-known mechanisms envisioned for use include silencing of the transgene with an siRNA or RNAi molecule (*e.g.*, miR, shRNA). For example, the ceDNA vector can comprise a regulatory switch that encodes a RNAi molecule that is complementary to the transgene expressed by the ceDNA vector. When such RNAi is expressed even if the transgene is expressed by the ceDNA vector, it will be silenced by the complementary RNAi molecule, and when the RNAi is not expressed when the transgene is expressed by the ceDNA vector the transgene is not silenced by the RNAi. Such an example of a RNAi molecule controlling gene expression, or as a regulatory switch is disclosed in US2017/0183664. In some embodiments, the regulatory switch comprises a repressor that blocks expression of the transgene from the ceDNA vector. In some embodiments, the on/off switch is a Small transcription activating RNA (STAR)-

based switch, for example, such as the one disclosed in Chappell J. *et al.,* Nat Chem Biol. 2015 Mar;11(3):214-20; and Chappell *et al.*, Microbiol Spectr. 2018 May;6(3. In some embodiments, the regulatory switch is a toehold switch, such as that disclosed in US2009/0191546, US2016/0076083, WO2017/087530, US2017/0204477, WO2017/075486 and in Green *et al,* Cell, 2014; 159(4); 925-939, all of which are incorporated by reference in their entireties herein.

[00301] In some embodiments, the regulatory switch is a tissue-specific self-inactivating regulatory switch, for example as disclosed in US2002/0022018, whereby the regulatory switch deliberately switches transgene expression off at a site where transgene expression might otherwise be disadvantageous. In some embodiments, the regulatory switch is a recombinase reversible gene expression system, for example as disclosed in US2014/0127162 and US Patent 8,324,436.

[00302] In some embodiments, the regulatory switch to control the transgene or gene of interest expressed by the ceDNA vector is a hybrid of a nucleic acid-based control mechanism and a small molecule regulator system. Such systems are well known to persons of ordinary skill in the art and are envisioned for use herein. Examples of such regulatory switches include, but are not limited to, an LTRi system or "Lac-Tet-RNAi" system, *e.g.*, as disclosed in US2010/0175141 and in Deans T. *et al.*, Cell., 2007, 130(2); 363-372, WO2008/051854 and US patent 9,388,425.

[00303] In some embodiments, the regulatory switch to control the transgene or gene of interest expressed by the ceDNA vector involves circular permutation, as disclosed in US patent 8,338,138. In such an embodiment, the molecular switch is multistable, *i.e.*, able to switch between at least two states, or alternatively, bistable, *i.e.*, a state is either "ON" or "OFF," for example, able to emit light or not, able to bind or not, able to catalyze or not, able to transfer electrons or not, and so forth. In another aspect, the molecular switch uses a fusion molecule, therefore the switch is able to switch between more than two states. For example, in response to a particular threshold state exhibited by an insertion sequence or acceptor sequence, the respective other sequence of the fusion may exhibit a range of states (*e.g.*, a range of binding activity, a range of enzyme catalysis, etc.). Thus, rather than switching from "ON" or "OFF," the fusion molecule can exhibit a graded response to a stimulus.

[00304] In some embodiments, a nucleic acid based regulatory switch can be selected from any or a combination of the switches in Table 11.

[00305] *E. Post-transcriptional and post-translational regulatory switches.*

[00306] In some embodiments, the regulatory switch to control the transgene or gene of interest expressed by the ceDNA vector is a post-transcriptional modification system. For example, such a regulatory switch can be an aptazyme riboswitch that is sensitive to tetracycline or theophylline, as disclosed in US2018/0119156, GB201107768, WO2001/064956A3, EP Patent 2707487 and Beilstein *et al.*, ACS Synth. Biol., 2015, 4 (5), pp 526–534; Zhong *et al.*, Elife. 2016 Nov 2;5. pii: e18858. In some embodiments, it is envisioned that a person of ordinary skill in the art could encode both the transgene and an inhibitory siRNA which contains a ligand sensitive (OFF-switch) aptamer, the net result being a ligand sensitive ON-switch.

[00307] In some embodiments, the regulatory switch to control the transgene or gene of interest expressed by the ceDNA vector is a post-translational modification system. In alternative embodiments, the gene of interest or protein is expressed as pro-protein or pre-proprotein, or has a signal response element (SRE) or a destabilizing domain (DD) attached to the expressed protein, thereby preventing correct protein folding and/or activity until post-translation modification has occurred. In the case of a destabilizing domain (DD) or SRE, the de-stabilization domain is post-translationally cleaved in the presence of an exogenous agent or small molecule. One of ordinary skill in the art can utilize such control methods as disclosed in US patent 8,173,792 and PCT application WO2017180587. Other post-transcriptional control switches envisioned for use in the ceDNA vector for controlling functional transgene activity are disclosed in Rakhit *et al.*, Chem Biol. 2014;21(9):1238-52 and Navarro *et al.*, ACS Chem Biol. 2016; 19; 11(8): 2101–2104A.

[00308] In some embodiments, a regulatory switch to control the transgene or gene of interest expressed by the ceDNA vector is a post-translational modification system that incorporates ligand sensitive inteins into the transgene coding sequence, such that the transgene or expressed protein is inhibited prior to splicing. For example, this has been demonstrated using both 4-hydroxytamoxifen and thyroid hormone (see, *e.g.*, US patents 7,541,450, 9,200,045; 7,192,739, Buskirk, et al, Proc Natl Acad Sci U S A. 2004 Jul 20; 101(29): 10505–10510; ACS Synth Biol. 2016 Dec 16; 5(12): 1475–1484; and 2005 Feb; 14(2): 523–532. In some embodiments, a post-transcriptional based regulatory switch can be selected from any or a combination of the switches in Table 11.

[00309] *F. Other exemplary regulatory switches*

[00310] Any known regulatory switch can be used in the ceDNA vector to control the gene expression of the transgene expressed by the ceDNA vector, including those triggered by environmental changes. Additional examples include, but are not limited to; the BOC method of Suzuki *et al.*, Scientific Reports 8; 10051 (2018); genetic code expansion and a non-physiologic amino acid; radiation-controlled or ultra-sound controlled on/off switches (see, *e.g.*, Scott S *et al.*, Gene Ther. 2000 Jul;7(13):1121-5; US patents 5,612,318; 5,571,797; 5,770,581; 5,817,636; and WO1999/025385A1. In some embodiments, the regulatory switch is controlled by an implantable system, *e.g.*, as disclosed in US patent 7,840,263; US2007/0190028A1 where gene expression is controlled by one or more forms of energy, including electromagnetic energy, that activates promoters operatively linked to the transgene in the ceDNA vector.

[00311] In some embodiments, a regulatory switch envisioned for use in the ceDNA vector is a hypoxia-mediated or stress-activated switch, *e.g.*, such as those disclosed in WO1999060142A2, US patent 5,834,306; 6,218,179; 6,709,858; US2015/0322410; Greco *et al.*, (2004) Targeted Cancer Therapies 9, S368, as well as FROG, TOAD and NRSE elements and conditionally inducible silence elements, including hypoxia response elements (HREs), inflammatory response elements (IREs) and shear-stress activated elements (SSAEs), *e.g.*,, as disclosed in U.S. Patent 9,394,526. Such an embodiment is useful for turning on expression of the transgene from the ceDNA vector after ischemia or in ischemic tissues, and/or tumors.

[00312] In some embodiments, a regulatory switch envisioned for use in the ceDNA vector is an optogenetic (*e.g.*, light controlled) regulatory switch, *e.g.*, such as one of the switches reviewed in Polesskaya *et al.*, BMC Neurosci. 2018; 19(Suppl 1): 12, and are also envisioned for use herein. In such embodiments, a ceDNA vector can comprise genetic elements are light sensitive and can regulate transgene expression in response to visible wavelengths (*e.g.* blue, near IR). ceDNA vectors comprising optogenetic regulatory switches are useful when expressing the transgene in locations of the body that can receive such light sources, *e.g.*, the skin, eye, muscle etc., and can also be used when ceDNA vectors are expressing transgenes in internal organs and tissues, where the light signal can be provided by a suitable means (*e.g.*, implantable device as disclosed herein). Such optogenetic regulatory switches include use of the light responsive elements, or light-inducible transcriptional effector (LITE) (*e.g.*, disclosed in 2014/0287938), a Light-On system (*e.g.*, disclosed in Wang *et al.*, Nat Methods. 2012 Feb 12;9(3):266-9; which has reported to enable *in vivo* control of expression of an insulin transgene, the Cry2/CIB1 system (*e.g.*, disclosed on Kennedy *et al.*, Nature Methods; 7, 973–975 (2010); and the FKF1/GIGANTEA system (*e.g.*, disclosed in Yazawa *et al.*, Nat Biotechnol. 2009 Oct;27(10):941-5).

[00313] *G. Kill Switches*

[00314] Other embodiments of the invention relate to a ceDNA vector comprising a kill switch. A kill switch as disclosed herein enables a cell comprising the ceDNA vector to be killed or undergo programmed cell death as a means to permanently remove an introduced ceDNA vector from the subject's system. It will be appreciated by one of ordinary skill in the art that use of kill switches in the ceDNA vectors of the invention would be typically coupled with targeting of the ceDNA vector to a limited number of cells that the subject can acceptably lose or to a cell type where apoptosis is desirable (*e.g.*, cancer cells). In all aspects, a "kill switch" as disclosed herein is designed to provide rapid and robust cell killing of the cell comprising the ceDNA vector in the absence of an input survival signal or other specified condition. Stated another way, a kill switch encoded by a ceDNA vector herein can restrict cell survival of a cell comprising a ceDNA vector to an environment defined by specific input signals. Such kill switches serve as a biological biocontainment function should it be desirable to remove the ceDNA vector from a subject or to ensure that it will not express the encoded transgene. Accordingly, kill switches are synthetic biological circuits in the ceDNA vector that couple environmental signals with conditional survival of the cell comprising the ceDNA vector. In some embodiments different ceDNA vectors can be designed to have different kill switches. This permits one to be able to control which transgene expressing cells are killed if cocktails of ceDNA vectors are used.

[00315] In some embodiments, a ceDNA vector can comprise a kill switch which is a modular biological containment circuit. In some embodiments, a kill switch encompassed for use in the ceDNA vector is disclosed in WO2017/059245, which describes a switch referred to as a "Deadman kill switch" that comprises a mutually inhibitory arrangement of at least two repressible sequences, such that an environmental signal represses the activity of a second molecule in the construct (*e.g.*, a small molecule-

binding transcription factor is used to produce a 'survival' state due to repression of toxin production). In cells comprising a ceDNA vector comprising a deadman kill switch, upon loss of the environmental signal, the circuit switches permanently to the 'death' state, where the toxin is now derepressed, resulting in toxin production which kills the cell. In another embodiment, a synthetic biological circuit referred to as a "Passcode circuit" or "Passcode kill switch" that uses hybrid transcription factors (TFs) to construct complex environmental requirements for cell survival, is provided. The Deadman and Passcode kill switches described in WO2017/059245 are particularly useful for use in ceDNA vectors, as they are modular and customizable, both in terms of the environmental conditions that control circuit activation and in the output modules that control cell fate. With the proper choice of toxins, including, but not limited to an endonuclease, *e.g.*, a EcoRI, Passcode circuits present in the ceDNA vector can be used to not only kill the host cell comprising the ceDNA vector, but also to degrade its genome and accompanying plasmids.

[00316] Other kill switches known to a person of ordinary skill in the art are encompassed for use in the ceDNA vector as disclosed herein, *e.g.*, as disclosed in US2010/0175141; US2013/0009799; US2011/0172826; US2013/0109568, as well as kill switches disclosed in Jusiak et al, Reviews in Cell Biology and molecular Medicine; 2014; 1-56; Kobayashi *et al.*, PNAS, 2004; 101; 8419-9; Marchisio *et al.*, Int. Journal of Biochem and Cell Biol., 2011; 43; 310-319; and in Reinshagen *et al.*, Science Translational Medicine, 2018, 11.

[00317] Accordingly, in some embodiments, the ceDNA vector can comprise a kill switch nucleic acid construct, which comprises the nucleic acid encoding an effector toxin or reporter protein, where the expression of the effector toxin (*e.g.*, a death protein) or reporter protein is controlled by a predetermined condition. For example, a predetermined condition can be the presence of an environmental agent, such as, *e.g.*, an exogenous agent, without which the cell will default to expression of the effector toxin (*e.g.*, a death protein) and be killed. In alternative embodiments, a predetermined condition is the presence of two or more environmental agents, *e.g.*, the cell will only survive when two or more necessary exogenous agents are supplied, and without either of which, the cell comprising the ceDNA vector is killed.

[00318] In some embodiments, the ceDNA vector is modified to incorporate a kill-switch to destroy the cells comprising the ceDNA vector to effectively terminate the *in vivo* expression of the transgene being expressed by the ceDNA vector (*e.g.*, therapeutic gene, protein or peptide etc). Specifically, the ceDNA vector is further genetically engineered to express a switch-protein that is not functional in mammalian cells under normal physiological conditions. Only upon administration of a drug or environmental condition that specifically targets this switch-protein, the cells expressing the switch-protein will be destroyed thereby terminating the expression of the therapeutic protein or peptide. For instance, it was reported that cells expressing HSV-thymidine kinase can be killed upon administration of drugs, such as ganciclovir and cytosine deaminase. See, for example, Dey and Evans, Suicide Gene Therapy by Herpes Simplex Virus-1 Thymidine Kinase (HSV-TK), in Targets in Gene Therapy, edited by You (2011); and Beltinger *et al.*, Proc. Natl. Acad. Sci. USA 96(15):8699-8704 (1999). In some embodiments the ceDNA vector can comprise a

siRNA kill switch referred to as DISE (Death Induced by Survival gene Elimination) (Murmann *et al.*, Oncotarget. 2017; 8:84643-84658. Induction of DISE in ovarian cancer cells in vivo).

[00319] In some aspects, a deadman kill switch is a biological circuit or system rendering a cellular response sensitive to a predetermined condition, such as the lack of an agent in the cell growth environment, *e.g.*, an exogenous agent. Such a circuit or system can comprise a nucleic acid construct comprising expression modules that form a deadman regulatory circuit sensitive to the predetermined condition, the construct comprising expression modules that form a regulatory circuit, the construct including:

i) a first repressor protein expression module, wherein the first repressor protein binds a first repressor protein nucleic acid binding element and represses transcription from a coding sequence comprising the first repressor protein binding element, and wherein repression activity of the first repressor protein is sensitive to inhibition by a first exogenous agent, the presence or absence of the first exogenous agent establishing a predetermined condition;

ii) a second repressor protein expression module, wherein the second repressor protein binds a second repressor protein nucleic acid binding element and represses transcription from a coding sequence comprising the second repressor protein binding element, wherein the second repressor protein is different from the first repressor protein; and

iii) an effector expression module, comprising a nucleic acid sequence encoding an effector protein, operably linked to a genetic element comprising a binding element for the second repressor protein, such that expression of the second repressor protein causes repression of effector expression from the effector expression module, wherein the second expression module comprises a first repressor protein nucleic acid binding element that permits repression of transcription of the second repressor protein when the element is bound by the first repressor protein, the respective modules forming a regulatory circuit such that in the absence of the first exogenous agent, the first repressor protein is produced from the first repressor protein expression module and represses transcription from the second repressor protein expression module, such that repression of effector expression by the second repressor protein is relieved, resulting in expression of the effector protein, but in the presence of the first exogenous agent, the activity of the first repressor protein is inhibited, permitting expression of the second repressor protein, which maintains expression of effector protein expression in the "off" state, such that the first exogenous agent is required by the circuit to maintain effector protein expression in the "off' state, and removal or absence of the first exogenous agent defaults to expression of the effector protein.

[00320] In some embodiments, the effector is a toxin or a protein that induces a cell death program. Any protein that is toxic to the host cell can be used. In some embodiments the toxin only kills those cells in which it is expressed. In other embodiments, the toxin kills other cells of the same host organism. Any of a large number of products that will lead to cell death can be employed in a deadman kill switch. Agents that inhibit DNA replication, protein translation or other processes or, *e.g.*, that degrade the host cell's nucleic acid, are of particular usefulness. To identify an efficient mechanism to kill the host cells upon circuit activation, several toxin genes were tested that directly damage the host cell's DNA or RNA. The

endonuclease *ecoRI*[21], the DNA gyrase inhibitor *ccdB*[22] and the ribonuclease-type toxin *mazF*[23] were tested because they are well-characterized, are native to *E. coli*, and provide a range of killing mechanisms. To increase the robustness of the circuit and provide an independent method of circuit-dependent cell death, the system can be further adapted to express, *e.g.*, a targeted protease or nuclease that further interferes with the repressor that maintains the death gene in the "off" state. Upon loss or withdrawal of the survival signal, death gene repression is even more efficiently removed by, *e.g.*, active degradation of the repressor protein or its message. As non-limiting examples, *mf*-Lon protease was used to not only degrade LacI but also target essential proteins for degradation. The *mf*-Lon degradation tag pdt#1 can be attached to the 3' end of five essential genes whose protein products are particularly sensitive to *mf*-Lon degradation[20], and cell viability was measured following removal of ATc. Among the tested essential gene targets, the peptidoglycan biosynthesis gene *murC* provided the strongest and fastest cell death phenotype (survival ratio $< 1 \times 10^{-4}$ within 6 hours).

[00321] As used herein, the term "predetermined input" refers to an agent or condition that influences the activity of a transcription factor polypeptide in a known manner. Generally, such agents can bind to and/or change the conformation of the transcription factor polypeptide to thereby modify the activity of the transcription factor polypeptide. Examples of predetermined inputs include, but are not limited to, environmental input agents that are not required for the survival of a given host organism (*i.e.*, in the absence of a synthetic biological circuit as described herein). Conditions that can provide a predetermined input include, for example temperature, *e.g.*, where the activity of one or more factors is temperature-sensitive, the presence or absence of light, including light of a given spectrum of wavelengths, and the concentration of a gas, salt, metal or mineral. Environmental input agents include, for example, a small molecule, biological agents such as pheromones, hormones, growth factors, metabolites, nutrients, and the like and analogs thereof; concentrations of chemicals, environmental byproducts, metal ions, and other such molecules or agents; light levels; temperature; mechanical stress or pressure; or electrical signals, such as currents and voltages.

[00322] In some embodiments, reporters are used to quantify the strength or activity of the signal received by the modules or programmable synthetic biological circuits of the invention. In some embodiments, reporters can be fused in-frame to other protein coding sequences to identify where a protein is located in a cell or organism. Luciferases can be used as effector proteins for various embodiments described herein, for example, measuring low levels of gene expression, because cells tend to have little to no background luminescence in the absence of a luciferase. In other embodiments, enzymes that produce colored substrates can be quantified using spectrophotometers or other instruments that can take absorbance measurements including plate readers. Like luciferases, enzymes like β-galactosidase can be used for measuring low levels of gene expression because they tend to amplify low signals. In some embodiments, an effector protein can be an enzyme that can degrade or otherwise destroy a given toxin. In some embodiments, an effector protein can be an odorant enzyme that converts a substrate to an odorant product. In some embodiments, an

effector protein can be an enzyme that phosphorylates or dephosphorylates either small molecules or other proteins, or an enzyme that methylates or demethylates other proteins or DNA.

[00323] In some embodiments, an effector protein can be a receptor, ligand, or lytic protein. Receptors tend to have three domains: an extracellular domain for binding ligands such as proteins, peptides or small molecules, a transmembrane domain, and an intracellular or cytoplasmic domain which frequently can participate in some sort of signal transduction event such as phosphorylation. In some embodiments, transporter, channel, or pump gene sequences are used as effector proteins. Non-limiting examples and sequences of effector proteins for use with the kill switches as described herein can be found at the Registry of Standard Biological Parts on the world wide web at parts.igem.org.

[00324] As used herein, a "modulator protein" is a protein that modulates the expression from a target nucleic acid sequence. Modulator proteins include, for example, transcription factors, including transcriptional activators and repressors, among others, and proteins that bind to or modify a transcription factor and influence its activity. In some embodiments, a modulator protein includes, for example, a protease that degrades a protein factor involved in the regulation of expression from a target nucleic acid sequence. Preferred modulator proteins include modular proteins in which, for example, DNA-binding and input agent-binding or responsive elements or domains are separable and transferrable, such that, for example, the fusion of the DNA binding domain of a first modulator protein to the input agent-responsive domain of a second results in a new protein that binds the DNA sequence recognized by the first protein, yet is sensitive to the input agent to which the second protein normally responds. Accordingly, as used herein, the term "modulator polypeptide," and the more specific "repressor polypeptide" include, in addition to the specified polypeptides, *e.g.*, "a LacI (repressor) polypeptide," variants, or derivatives of such polypeptides that responds to a different or variant input agent. Thus, for a LacI polypeptide, included are LacI mutants or variants that bind to agents other than lactose or IPTG. A wide range of such agents are known in the art.

[00325] Table 11. Exemplary regulatory switches. [b]ON switchability by an effector; other than removing the effector which confers the OFF state. [c]OFF switchability by an effector; other than removing the effector which confers the ON state. [d]A ligand or other physical stimuli (*e.g.* temperature, electromagnetic radiation, electricity) which stabilizes the switch either in its ON or OFF state. [e]refers to the reference number cited in Kis *et al.,* J R Soc Interface. 12:20141000 (2015), where both the article and the references cited therein are hereby incorporated by reference herein.

no.	name	ON switch[b]	OFF switch[c]	origin	effector[d]	references[e]
Transcriptional Switches						
1	ABA	yes	no	*Arabidopsis thaliana*, yeast	abscisic acid	[19]
2	AIR	yes	no	*Aspergillus nidulans*	acetaldehyde	[20]

3	ART	yes	no	*Chlamydia pneumoniae*	l-arginine	[21]
4	BEARON, BEAROFF	yes	yes	*Campylobacter jejuni*	bile acid	[22]
5	BirA-tTA	no	yes	*Escherichia coli*	biotin (vitamin H)	[23]
6	BIT	yes	no	*Escherichia coli*	biotin (vitamin H)	[24]
7	Cry2-CIB1	yes	no	*Arabidopsis thaliana*, yeast	blue light	[25]
8	CTA, CTS	yes	yes	*Comamonas testosteroni*, *Homo sapiens*	food additives (benzoate, vanillate)	[26]
9	cTA, rcTA	yes	yes	*Pseudomonas putida*	cumate	[27]
10	Ecdysone	yes	no	*Homo sapiens, Drosophila melanogaster*	Ecdysone	[28]
11	EcR:RXR	yes	no	*Homo sapiens*, *Locusta migratoria*	ecdysone	[29]
12	electro-genetic	yes	no	*Aspergillus nidulans*	electricity, acetaldehyde	[30]
13	ER-p65-ZF	yes	no	*Homo sapiens*, yeast	4,4′-dyhydroxybenzil	[31]
14	E.REX	yes	yes	*Escherichia coli*	erythromycin	[32]
15	EthR	no	yes	*Mycobacterium tuberculosis*	2-phenylethyl-butyrate	[33]
16	GAL4-ER	yes	yes	yeast, *Homo sapiens*	oestrogen, 4-hydroxytamoxifen	[34]
17	GAL4-hPR	yes	yes	yeast, *Homo sapiens*	mifepristone	[35,36]
18	GAL4-Raps	yes	yes	yeast, *Homo sapiens*	rapamycin and rapamycin derivatives	[37]
19	GAL4-TR	yes	no	yeast, *Homo sapiens*	thyroid hormone	[38]

20	GyrB	yes	yes	*Escherichia coli*	coumermycin, novobiocin	[39]
21	HEA-3	yes	no	*Homo sapiens*	4-hydroxytamoxifen	[40]
22	Intramer	no	yes	synthetic SELEX-derived aptamers	theophylline	[41]
23	LacI	yes	no	*Escherichia coli*	IPTG	[42–46]
24	LAD	yes	no	*Arabidopsis thaliana*, yeast	blue light	[47]
25	LightOn	yes	no	*Neurospora crassa*, yeast	blue light	[48]
26	NICE	yes	yes	*Arthrobacter nicotinovorans*	6-hydroxynicotine	[49]
27	PPAR*	yes	no	*Homo sapiens*	rosiglitazone	[50]
28	PEACE	no	yes	*Pseudomonas putida*	flavonoids (*e.g.* phloretin)	[51]
29	PIT	yes	yes	*Streptomyces coelicolor*	pristinamycin I, virginiamycin	[12]
30	REDOX	no	yes	*Streptomyces coelicolor*	NADH	[52]
31	QuoRex	yes	yes	*Streptomyces coelicolor*, *Streptomyces pristinaespiralis*	butyrolactones (*e.g.* SCB1)	[53]
32	ST-TA	yes	yes	*Streptomyces coelicolor*, *Escherichia coli*, *Herpes simplex*	γ-butyrolactone, tetracycline	[54]
33	TIGR	no	yes	*Streptomyces albus*	temperature	[55]
34	TraR	yes	no	*Agrobacterium tumefaciens*	*N*-(3-oxo-octanoyl)homoserine lactone	[56]

35	TET-OFF, TET-ON	yes	yes	*Escherichia coli*, *Herpes simplex*	tetracycline, doxycycline	[11,57]
36	TRT	yes	no	*Chlamydia trachomatis*	l-tryptophan	[58]
37	UREX	yes	no	*Deinococcus radiodurans*	uric acid	[59]
38	VAC	yes	yes	*Caulobacter crescentus*	vanillic acid	[60]
39	ZF-ER, ZF-RXR/EcR	yes	yes	*Mus musculus*, *Homo sapiens*, *Drosophila melanogaster*	4-hydroxytamoxifen, ponasterone-A	[61]
40	ZF-Raps	yes	no	*Homo sapiens*	rapamycin	[62]
41	ZF switches	yes	no	*Mus musculus*, *Homo sapiens*, *Drosophila melanogaster*	4-hydroxytamoxifen, mifepristone	[63]
42	ZF(TF)s	yes	no	*Xenopus laevis*, *Homo sapiens*	ethyl-4-hydroxybenzoate, propyl-4-hydroxybenzoate	[64]
post-transcriptional switches						
1	aptamer RNAi	yes	no	synthetic SELEX-derived aptamer	theophylline	[65]
2	aptamer RNAi	no	yes	synthetic SELEX-derived aptamer	theophylline	[66]
3	aptamer RNAi miRNA	yes	no	synthetic SELEX-derived aptamer	theophylline, tetracycline, hypoxanthine	[67]
4	aptamer Splicing	yes	yes	*Homo sapiens*, MS2 bacteriophage	MS2, p65, p50, b-catenin	[68]

5	aptazyme	no	yes	synthetic SELEX-derived aptamer, *Schistosoma mansoni*	theophylline	[69]
6	replicon CytTS	yes	no	Sindbis virus	temperature	[70]
7	TET-OFF-shRNA, TET-ON-shRNA	yes	yes	*Escherichia coli*, *Herpes simplex*, *Homo sapiens*	doxycycline	[71]
8	theo aptamer	no	yes	synthetic SELEX-derived aptamer	theophylline	[72]
9	3′ UTR aptazyme	yes	no	synthetic SELEX-derived aptamers, tobacco ringspot virus	theophylline, tetracycline	[73]
10	5′ UTR aptazyme	no	yes	synthetic SELEX-derived aptamer, *Schistosoma mansoni*	theophylline	[74]
translational switches						
1	Hoechst aptamer	no	yes	synthetic RNA sequence	Hoechst dyes	[75]
2	H23 aptamer	no	yes	*Archaeoglobus fulgidus*	L7Ae, L7KK	[76]
3	L7Ae aptamer	yes	yes	*Archaeoglobus fulgidus*	L7Ae	[77]
4	MS2 aptamer	no	yes	MS2 bacteriophage	MS2	[78]
post-translational switches						
1	AID	no	yes	*Arabidopsis thaliana*, *Oryza sativa*, *Gossypium hirsutum*	auxins (*e.g.* IAA)	[79]

2	ER DD	no	yes	*Homo sapiens*	CMP8, 4-hydroxytamoxifen	[80]
3	FM	yes	no	*Homo sapiens*	AP21998	[81]
4	HaloTag	no	yes	*Rhodococcus* sp. RHA1	HyT13	[82,83]
5	HDV-aptazyme	no	yes	hepatitis delta virus	theophylline, guanine	[84]
6	PROTAC	no	yes	*Homo sapiens*	proteolysis targeting chimeric molecules (PROTACS)	[85]
7	shield DD	yes	no	*Homo sapiens*	shields (*e.g.* Shld1)	[86]
8	shield LID	no	yes	*Homo sapiens*	shields (*e.g.* Shld1)	[87]
9	TMP DD	yes	no	*Escherichia coli*	trimethoprim (TMP)	[88]

VII. Pharmaceutical Compositions

[00326] In another aspect, pharmaceutical compositions are provided. The pharmaceutical composition comprises a ceDNA vector as disclosed herein and a pharmaceutically acceptable carrier or diluent.

[00327] The DNA-vectors disclosed herein can be incorporated into pharmaceutical compositions suitable for administration to a subject for *in vivo* delivery to cells, tissues, or organs of the subject. Typically, the pharmaceutical composition comprises a ceDNA-vector as disclosed herein and a pharmaceutically acceptable carrier. For example, the ceDNA vectors described herein can be incorporated into a pharmaceutical composition suitable for a desired route of therapeutic administration (*e.g.*, parenteral administration). Passive tissue transduction via high pressure intravenous or intraarterial infusion, as well as intracellular injection, such as intranuclear microinjection or intracytoplasmic injection, are also contemplated. Pharmaceutical compositions for therapeutic purposes can be formulated as a solution, microemulsion, dispersion, liposomes, or other ordered structure suitable to high ceDNA vector concentration. Sterile injectable solutions can be prepared by incorporating the ceDNA vector compound in the required amount in an appropriate buffer with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization.

[00328] Pharmaceutically active compositions comprising a ceDNA vector can be formulated to deliver a transgene in the nucleic acid to the cells of a recipient, resulting in the therapeutic expression of the transgene therein. The composition can also include a pharmaceutically acceptable carrier.

[00329] A ceDNA vector as disclosed herein can be incorporated into a pharmaceutical composition suitable for topical, systemic, intra-amniotic, intrathecal, intracranial, intraarterial, intravenous, intralymphatic, intraperitoneal, subcutaneous, tracheal, intra-tissue (*e.g.*, intramuscular, intracardiac, intrahepatic, intrarenal, intracerebral), intrathecal, intravesical, conjunctival (*e.g.*, extra-orbital, intraorbital, retroorbital, intraretinal, subretinal, choroidal, sub-choroidal, intrastromal, intracameral and intravitreal), intracochlear, and mucosal (*e.g.*, oral, rectal, nasal) administration. Passive tissue transduction via high pressure intravenous or intraarterial infusion, as well as intracellular injection, such as intranuclear microinjection or intracytoplasmic injection, are also contemplated.

[00330] Pharmaceutical compositions for therapeutic purposes typically must be sterile and stable under the conditions of manufacture and storage. The composition can be formulated as a solution, microemulsion, dispersion, liposomes, or other ordered structure suitable to high ceDNA vector concentration. Sterile injectable solutions can be prepared by incorporating the ceDNA vector compound in the required amount in an appropriate buffer with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization.

[00331] Various techniques and methods are known in the art for delivering nucleic acids to cells. For example, nucleic acids, such as ceDNA can be formulated into lipid nanoparticles (LNPs), lipidoids, liposomes, lipid nanoparticles, lipoplexes, or core-shell nanoparticles. Typically, LNPs are composed of nucleic acid (*e.g.*, ceDNA) molecules, one or more ionizable or cationic lipids (or salts thereof), one or more non-ionic or neutral lipids (*e.g.*, a phospholipid), a molecule that prevents aggregation (*e.g.*, PEG or a PEG-lipid conjugate), and optionally a sterol (*e.g.*, cholesterol).

[00332] Another method for delivering nucleic acids, such as ceDNA to a cell is by conjugating the nucleic acid with a ligand that is internalized by the cell. For example, the ligand can bind a receptor on the cell surface and internalized via endocytosis. The ligand can be covalently linked to a nucleotide in the nucleic acid. Exemplary conjugates for delivering nucleic acids into a cell are described, example, in WO2015/006740, WO2014/025805, WO2012/037254, WO2009/082606, WO2009/073809, WO2009/018332, WO2006/112872, WO2004/090108, WO2004/091515 and WO2017/177326.

[00333] Nucleic acids, such as ceDNA, can also be delivered to a cell by transfection. Useful transfection methods include, but are not limited to, lipid-mediated transfection, cationic polymer-mediated transfection, or calcium phosphate precipitation. Transfection reagents are well known in the art and include, but are not limited to, TurboFect Transfection Reagent (Thermo Fisher Scientific), Pro-Ject Reagent (Thermo Fisher Scientific), TRANSPASS™ P Protein Transfection Reagent (New England Biolabs), CHARIOT™ Protein Delivery Reagent (Active Motif), PROTEOJUICE™ Protein Transfection Reagent (EMD Millipore), 293fectin, LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ 3000 (Thermo Fisher Scientific), LIPOFECTAMINE™ (Thermo Fisher Scientific), LIPOFECTIN™ (Thermo Fisher Scientific), DMRIE-C, CELLFECTIN™ (Thermo Fisher Scientific), OLIGOFECTAMINE™ (Thermo Fisher Scientific), LIPOFECTACE™, FUGENE™ (Roche, Basel, Switzerland), FUGENE™ HD (Roche),

TRANSFECTAM™(Transfectam, Promega, Madison, Wis.), TFX-10™ (Promega), TFX-20™ (Promega), TFX-50™ (Promega), TRANSFECTIN™ (BioRad, Hercules, Calif.), SILENTFECT™ (Bio-Rad), Effectene™ (Qiagen, Valencia, Calif.), DC-chol (Avanti Polar Lipids), GENEPORTER™ (Gene Therapy Systems, San Diego, Calif.), DHARMAFECT 1™ (Dharmacon, Lafayette, Colo.), DHARMAFECT 2™ (Dharmacon), DHARMAFECT 3™ (Dharmacon), DHARMAFECT 4™ (Dharmacon), ESCORT™ III (Sigma, St. Louis, Mo.), and ESCORT™ IV (Sigma Chemical Co.). Nucleic acids, such as ceDNA, can also be delivered to a cell via microfluidics methods known to those of skill in the art.

[00334] Methods of non-viral delivery of nucleic acids *in vivo* or *ex vivo* include electroporation, lipofection (see, U.S. Pat. No. 5,049,386; 4,946,787 and commercially available reagents such as Transfectam™ and Lipofectin™), microinjection, biolistics, virosomes, liposomes (see, *e.g.*, Crystal, Science 270:404-410 (1995); Blaese *et al.*, Cancer Gene Ther. 2:291-297 (1995); Behr *et al.*, Bioconjugate Chem. 5:382-389 (1994); Remy *et al.*, Bioconjugate Chem. 5:647-654 (1994); Gao *et al.*, Gene Therapy 2:710-722 (1995); Ahmad *et al.*, Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787), immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, and agent-enhanced uptake of DNA. Sonoporation using, *e.g.*, the Sonitron 2000 system (Rich-Mar) can also be used for delivery of nucleic acids.

[00335] ceDNA vectors as described herein can also be administered directly to an organism for transduction of cells *in vivo*. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells including, but not limited to, injection, infusion, topical application and electroporation. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[00336] Methods for introduction of a nucleic acid vector ceDNA vector as disclosed herein can be delivered into hematopoietic stem cells, for example, by the methods as described, for example, in U.S. Pat. No. 5,928,638.

[00337] The ceDNA vectors in accordance with the present invention can be added to liposomes for delivery to a cell or target organ in a subject. Liposomes are vesicles that possess at least one lipid bilayer. Liposomes are typical used as carriers for drug/ therapeutic delivery in the context of pharmaceutical development. They work by fusing with a cellular membrane and repositioning its lipid structure to deliver a drug or active pharmaceutical ingredient (API). Liposome compositions for such delivery are composed of phospholipids, especially compounds having a phosphatidylcholine group, however these compositions may also include other lipids.

[00338] In some aspects, the disclosure provides for a liposome formulation that includes one or more compounds with a polyethylene glycol (PEG) functional group (so-called "PEG-ylated compounds") which can reduce the immunogenicity/ antigenicity of, provide hydrophilicity and hydrophobicity to the

compound(s) and reduce dosage frequency. Or the liposome formulation simply includes polyethylene glycol (PEG) polymer as an additional component. In such aspects, the molecular weight of the PEG or PEG functional group can be from 62 Da to about 5,000 Da.

[00339] In some aspects, the disclosure provides for a liposome formulation that will deliver an API with extended release or controlled release profile over a period of hours to weeks. In some related aspects, the liposome formulation may comprise aqueous chambers that are bound by lipid bilayers. In other related aspects, the liposome formulation encapsulates an API with components that undergo a physical transition at elevated temperature which releases the API over a period of hours to weeks.

[00340] In some aspects, the liposome formulation comprises sphingomyelin and one or more lipids disclosed herein. In some aspects, the liposome formulation comprises optisomes.

[00341] In some aspects, the disclosure provides for a liposome formulation that includes one or more lipids selected from: N-(carbonyl-methoxypolyethylene glycol 2000)-1,2-distearoyl-sn-glycero-3-phosphoethanolamine sodium salt, (distearoyl-sn-glycero-phosphoethanolamine), MPEG (methoxy polyethylene glycol)-conjugated lipid, HSPC (hydrogenated soy phosphatidylcholine); PEG (polyethylene glycol); DSPE (distearoyl-sn-glycero-phosphoethanolamine); DSPC (distearoylphosphatidylcholine); DOPC (dioleoylphosphatidylcholine); DPPG (dipalmitoylphosphatidylglycerol); EPC (egg phosphatidylcholine); DOPS (dioleoylphosphatidylserine); POPC (palmitoyloleoylphosphatidylcholine); SM (sphingomyelin); MPEG (methoxy polyethylene glycol); DMPC (dimyristoyl phosphatidylcholine); DMPG (dimyristoyl phosphatidylglycerol); DSPG (distearoylphosphatidylglycerol); DEPC (dierucoylphosphatidylcholine); DOPE (dioleoly-sn-glycero-phophoethanolamine). cholesteryl sulphate (CS), dipalmitoylphosphatidylglycerol (DPPG), DOPC (dioleoly-sn-glycero-phosphatidylcholine) or any combination thereof.

[00342] In some aspects, the disclosure provides for a liposome formulation comprising phospholipid, cholesterol and a PEG-ylated lipid in a molar ratio of 56:38:5. In some aspects, the liposome formulation's overall lipid content is from 2-16 mg/mL. In some aspects, the disclosure provides for a liposome formulation comprising a lipid containing a phosphatidylcholine functional group, a lipid containing an ethanolamine functional group and a PEG-ylated lipid. In some aspects, the disclosure provides for a liposome formulation comprising a lipid containing a phosphatidylcholine functional group, a lipid containing an ethanolamine functional group and a PEG-ylated lipid in a molar ratio of 3:0.015:2 respectively. In some aspects, the disclosure provides for a liposome formulation comprising a lipid containing a phosphatidylcholine functional group, cholesterol and a PEG-ylated lipid. In some aspects, the disclosure provides for a liposome formulation comprising a lipid containing a phosphatidylcholine functional group and cholesterol. In some aspects, the PEG-ylated lipid is PEG-2000-DSPE. In some aspects, the disclosure provides for a liposome formulation comprising DPPG, soy PC, MPEG-DSPE lipid conjugate and cholesterol.

[00343] In some aspects, the disclosure provides for a liposome formulation comprising one or more lipids containing a phosphatidylcholine functional group and one or more lipids containing an ethanolamine functional group. In some aspects, the disclosure provides for a liposome formulation comprising one or more: lipids containing a phosphatidylcholine functional group, lipids containing an ethanolamine functional group, and sterols, *e.g.* cholesterol. In some aspects, the liposome formulation comprises DOPC/ DEPC; and DOPE.

[00344] In some aspects, the disclosure provides for a liposome formulation further comprising one or more pharmaceutical excipients, *e.g.* sucrose and/or glycine.

[00345] In some aspects, the disclosure provides for a liposome formulation that is wither unilamellar or multilamellar in structure. In some aspects, the disclosure provides for a liposome formulation that comprises multi-vesicular particles and/or foam-based particles. In some aspects, the disclosure provides for a liposome formulation that are larger in relative size to common nanoparticles and about 150 to 250 nm in size. In some aspects, the liposome formulation is a lyophilized powder.

[00346] In some aspects, the disclosure provides for a liposome formulation that is made and loaded with ceDNA vectors disclosed or described herein, by adding a weak base to a mixture having the isolated ceDNA outside the liposome. This addition increases the pH outside the liposomes to approximately 7.3 and drives the API into the liposome. In some aspects, the disclosure provides for a liposome formulation having a pH that is acidic on the inside of the liposome. In such cases the inside of the liposome can be at pH 4-6.9, and more preferably pH 6.5. In other aspects, the disclosure provides for a liposome formulation made by using intra-liposomal drug stabilization technology. In such cases, polymeric or non-polymeric highly charged anions and intra-liposomal trapping agents are utilized, *e.g.* polyphosphate or sucrose octasulfate.

[00347] In other aspects, the disclosure provides for a liposome formulation comprising phospholipids, lecithin, phosphatidylcholine and phosphatidylethanolamine.

[00348] Delivery reagents such as liposomes, nanocapsules, microparticles, microspheres, lipid particles, vesicles, and the like, can be used for the introduction of the compositions of the present disclosure into suitable host cells. In particular, the nucleic acids can be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, a nanoparticle, a gold particle, or the like. Such formulations can be preferred for the introduction of pharmaceutically acceptable formulations of the nucleic acids disclosed herein.

[00349] Various delivery methods known in the art or modification thereof can be used to deliver ceDNA vectors *in vitro* or *in vivo*. For example, in some embodiments, ceDNA vectors are delivered by making transient penetration in cell membrane by mechanical, electrical, ultrasonic, hydrodynamic, or laser-based energy so that DNA entrance into the targeted cells is facilitated. For example, a ceDNA vector can be delivered by transiently disrupting cell membrane by squeezing the cell through a size-restricted channel or

by other means known in the art. In some cases, a ceDNA vector alone is directly injected as naked DNA into skin, thymus, cardiac muscle, skeletal muscle, or liver cells.

[00350] In some cases, a ceDNA vector is delivered by gene gun. Gold or tungsten spherical particles (1–3 μm diameter) coated with capsid-free AAV vectors can be accelerated to high speed by pressurized gas to penetrate into target tissue cells.

[00351] In some embodiments, electroporation is used to deliver ceDNA vectors. Electroporation causes temporary destabilization of the cell membrane target cell tissue by insertion of a pair of electrodes into the tissue so that DNA molecules in the surrounding media of the destabilized membrane would be able to penetrate into cytoplasm and nucleoplasm of the cell. Electroporation has been used *in vivo* for many types of tissues, such as skin, lung, and muscle.

[00352] In some cases, a ceDNA vector is delivered by hydrodynamic injection, which is a simple and highly efficient method for direct intracellular delivery of any water-soluble compounds and particles into internal organs and skeletal muscle in an entire limb.

[00353] In some cases, ceDNA vectors are delivered by ultrasound by making nanoscopic pores in membrane to facilitate intracellular delivery of DNA particles into cells of internal organs or tumors, so the size and concentration of plasmid DNA have great role in efficiency of the system. In some cases, ceDNA vectors are delivered by magnetofection by using magnetic fields to concentrate particles containing nucleic acid into the target cells.

[00354] In some cases, chemical delivery systems can be used, for example, by using nanomeric complexes, which include compaction of negatively charged nucleic acid by polycationic nanomeric particles, belonging to cationic liposome/micelle or cationic polymers. Cationic lipids used for the delivery method includes, but not limited to monovalent cationic lipids, polyvalent cationic lipids, guanidine containing compounds, cholesterol derivative compounds, cationic polymers, (*e.g.,* poly(ethylenimine), poly-L-lysine, protamine, other cationic polymers), and lipid-polymer hybrid.

[00355] A. Exosomes:

[00356] In some embodiments, a ceDNA vector as disclosed herein is delivered by being packaged in an exosome. Exosomes are small membrane vesicles of endocytic origin that are released into the extracellular environment following fusion of multivesicular bodies with the plasma membrane. Their surface consists of a lipid bilayer from the donor cell's cell membrane, they contain cytosol from the cell that produced the exosome, and exhibit membrane proteins from the parental cell on the surface. Exosomes are produced by various cell types including epithelial cells, B and T lymphocytes, mast cells (MC) as well as dendritic cells (DC). Some embodiments, exosomes with a diameter between 10nm and 1μm, between 20nm and 500nm, between 30nm and 250nm, between 50nm and 100nm are envisioned for use. Exosomes can be isolated for a delivery to target cells using either their donor cells or by introducing specific nucleic acids into them. Various approaches known in the art can be used to produce exosomes containing capsid-free AAV vectors of the present invention.

[00357] B. Microparticle/Nanoparticles:

[00358] In some embodiments, a ceDNA vector as disclosed herein is delivered by a lipid nanoparticle. Generally, lipid nanoparticles comprise an ionizable amino lipid (*e.g.*, heptatriaconta-6,9,28,31-tetraen-19-yl 4-(dimethylamino)butanoate, DLin-MC3-DMA, a phosphatidylcholine (1,2-distearoyl-sn-glycero-3-phosphocholine, DSPC), cholesterol and a coat lipid (polyethylene glycol-dimyristolglycerol, PEG-DMG), for example as disclosed by Tam *et al.* (2013). *Advances in Lipid Nanoparticles for siRNA delivery*. Pharmaceuticals 5(3): 498-507.

[00359] In some embodiments, a lipid nanoparticle has a mean diameter between about 10 and about 1000 nm. In some embodiments, a lipid nanoparticle has a diameter that is less than 300 nm. In some embodiments, a lipid nanoparticle has a diameter between about 10 and about 300 nm. In some embodiments, a lipid nanoparticle has a diameter that is less than 200 nm. In some embodiments, a lipid nanoparticle has a diameter between about 25 and about 200 nm. In some embodiments, a lipid nanoparticle preparation (*e.g.*, composition comprising a plurality of lipid nanoparticles) has a size distribution in which the mean size (*e.g.*, diameter) is about 70 nm to about 200 nm, and more typically the mean size is about 100 nm or less.

[00360] Various lipid nanoparticles known in the art can be used to deliver ceDNA vector disclosed herein. For example, various delivery methods using lipid nanoparticles are described in U.S. Patent Nos. 9,404,127, 9,006,417 and 9,518,272.

[00361] In some embodiments, a ceDNA vector disclosed herein is delivered by a gold nanoparticle. Generally, a nucleic acid can be covalently bound to a gold nanoparticle or non-covalently bound to a gold nanoparticle (*e.g.*, bound by a charge-charge interaction), for example as described by Ding *et al.* (2014). *Gold Nanoparticles for Nucleic Acid Delivery*. Mol. Ther. 22(6); 1075-1083. In some embodiments, gold nanoparticle-nucleic acid conjugates are produced using methods described, for example, in U.S. Patent No. 6,812,334.

[00362] C. Liposomes

[00363] The formation and use of liposomes is generally known to those of skill in the art. Liposomes have been developed with improved serum stability and circulation half-times (U.S. Pat. No. 5,741,516). Further, various methods of liposome and liposome like preparations as potential drug carriers have been described (U.S. Pat. Nos. 5,567,434; 5,552,157; 5,565,213; 5,738,868 and 5,795,587).

[00364] Liposomes have been used successfully with a number of cell types that are normally resistant to transfection by other procedures. In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, drugs, radiotherapeutic agents, viruses, transcription factors and allosteric effectors into a variety of cultured cell lines and animals. In addition, several successful clinical trials examining the effectiveness of liposome-mediated drug delivery have been completed.

[00365] Liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs). MLVs generally have diameters of from 25 nm to 4 μm. Sonication of MLVs results in the formation of small unilamellar vesicles (SUVs) with diameters in the range of 200 to 500 ANG., containing an aqueous solution in the core.

[00366] In some embodiments, a liposome comprises cationic lipids. The term "cationic lipid" includes lipids and synthetic lipids having both polar and non-polar domains and which are capable of being positively charged at or around physiological pH and which bind to polyanions, such as nucleic acids, and facilitate the delivery of nucleic acids into cells. In some embodiments, cationic lipids include saturated and unsaturated alkyl and alicyclic ethers and esters of amines, amides, or derivatives thereof. In some embodiments, cationic lipids comprise straight-chain, branched alkyl, alkenyl groups, or any combination of the foregoing. In some embodiments, cationic lipids contain from 1 to about 25 carbon atoms (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 carbon atoms. In some embodiments, cationic lipids contain more than 25 carbon atoms. In some embodiments, straight chain or branched alkyl or alkene groups have six or more carbon atoms. A cationic lipid can also comprise, in some embodiments, one or more alicyclic groups. Non-limiting examples of alicyclic groups include cholesterol and other steroid groups. In some embodiments, cationic lipids are prepared with a one or more counterions. Examples of counterions (anions) include but are not limited to Cl^-, Br^-, I^-, F^-, acetate, trifluoroacetate, sulfate, nitrite, and nitrate.

[00367] Non-limiting examples of cationic lipids include polyethylenimine, polyamidoamine (PAMAM) starburst dendrimers, Lipofectin (a combination of DOTMA and DOPE), Lipofectase, LIPOFECTAMINE™ (*e.g.*, LIPOFECTAMINE™ 2000), DOPE, Cytofectin (Gilead Sciences, Foster City, Calif.), and Eufectins (JBL, San Luis Obispo, Calif.). Exemplary cationic liposomes can be made from N-[1-(2,3-dioleoloxy)-propyl]-N,N,N-trimethylammonium chloride (DOTMA), N-[1 -(2,3-dioleoloxy)-propyl]-N,N,N-trimethylammonium methylsulfate (DOTAP), 3β-[N-(N′,N′-dimethylaminoethane)carbamoyl]cholesterol (DC-Chol), 2,3,-dioleyloxy-N-[2(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate (DOSPA), 1,2-dimyristyloxypropyl-3-dimethyl-hydroxyethyl ammonium bromide; and dimethyldioctadecylammonium bromide (DDAB). Nucleic acids (*e.g.*, CELiD) can also be complexed with, *e.g.*, poly (L-lysine) or avidin and lipids can, or cannot, be included in this mixture, *e.g.*, steryl-poly (L-lysine).

[00368] In some embodiments, a ceDNA vector as disclosed herein is delivered using a cationic lipid described in U.S. Patent No. 8,158,601, or a polyamine compound or lipid as described in U.S. Patent No. 8,034,376.

[00369] D. Conjugates

[00370] In some embodiments, a ceDNA vector as disclosed herein is conjugated (*e.g.*, covalently bound to an agent that increases cellular uptake. An "agent that increases cellular uptake" is a molecule that

facilitates transport of a nucleic acid across a lipid membrane. For example, a nucleic acid can be conjugated to a lipophilic compound (*e.g.*, cholesterol, tocopherol, *etc.*), a cell penetrating peptide (CPP) (*e.g.*, penetratin, TAT, Syn1B, *etc.*), and polyamines (*e.g.*, spermine). Further examples of agents that increase cellular uptake are disclosed, for example, in Winkler (2013). *Oligonucleotide conjugates for therapeutic applications*. Ther. Deliv. 4(7); 791-809.

[00371] In some embodiments, a ceDNA vector as disclosed herein is conjugated to a polymer (*e.g.*, a polymeric molecule) or a folate molecule (*e.g.*, folic acid molecule). Generally, delivery of nucleic acids conjugated to polymers is known in the art, for example as described in WO2000/34343 and WO2008/022309. In some embodiments, a ceDNA vector as disclosed herein is conjugated to a poly(amide) polymer, for example as described by U.S. Patent No. 8,987,377. In some embodiments, a nucleic acid described by the disclosure is conjugated to a folic acid molecule as described in U.S. Patent No. 8,507,455.

[00372] In some embodiments, a ceDNA vector as disclosed herein is conjugated to a carbohydrate, for example as described in U.S. Patent No. 8,450,467.

[00373] E. Nanocapsule

[00374] Alternatively, nanocapsule formulations of a ceDNA vector as disclosed herein can be used. Nanocapsules can generally entrap substances in a stable and reproducible way. To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) should be designed using polymers able to be degraded *in vivo*. Biodegradable polyalkyl-cyanoacrylate nanoparticles that meet these requirements are contemplated for use.

VIII. Methods of delivering ceDNA vectors

[00375] In some embodiments, a ceDNA vector can be delivered to a target cell *in vitro* or *in vivo* by various suitable methods. ceDNA vectors alone can be applied or injected. CeDNA vectors can be delivered to a cell without the help of a transfection reagent or other physical means. Alternatively, ceDNA vectors can be delivered using any art-known transfection reagent or other art-known physical means that facilitates entry of DNA into a cell, *e.g.*, liposomes, alcohols, polylysine- rich compounds, arginine-rich compounds, calcium phosphate, microvesicles, microinjection, electroporation and the like.

[00376] In contrast, transductions with capsid-free AAV vectors disclosed herein can efficiently target cell and tissue-types that are difficult to transduce with conventional AAV virions using various delivery reagent.

IX. Additional uses of the ceDNA vectors

[00377] The compositions and ceDNA vectors provided herein can be used to deliver a transgene for various purposes. In some embodiments, the transgene encodes a protein or functional RNA that is intended to be used for research purposes, *e.g.*, to create a somatic transgenic animal model harboring the transgene, *e.g.*, to study the function of the transgene product. In another example, the transgene encodes a protein or functional RNA that is intended to be used to create an animal model of disease. In some

embodiments, the transgene encodes one or more peptides, polypeptides, or proteins, which are useful for the treatment, prevention, or amelioration of disease states or disorders in a mammalian subject. The transgene can be transferred (*e.g.*, expressed in) to a subject in a sufficient amount to treat a disease associated with reduced expression, lack of expression or dysfunction of the gene. In some embodiments the transgene can be transferred to (*e.g.*, expressed in) a subject in a sufficient amount to treat a disease associated with increased expression, activity of the gene product, or inappropriate upregulation of a gene that the transgene suppresses or otherwise causes the expression of which to be reduced.

X. Methods of Use

[00378] The ceDNA vector of the invention can also be used in a method for the delivery of a nucleotide sequence of interest to a target cell. The method may in particular be a method for delivering a therapeutic gene of interest to a cell of a subject in need thereof. The invention allows for the in vivo expression of a polypeptide, protein, or oligonucleotide encoded by a therapeutic exogenous DNA sequence in cells in a subject such that therapeutic levels of the polypeptide, protein, or oligonucleotide are expressed. These results are seen with both in vivo and in vitro modes of ceDNA vector delivery.

[00379] A method for the delivery of a nucleic acid of interest in a cell of a subject can comprise the administration to said subject of a ceDNA vector of the invention comprising said nucleic acid of interest. In addition, the invention provides a method for the delivery of a nucleic acid of interest in a cell of a subject in need thereof, comprising multiple administrations of the ceDNA vector of the invention comprising said nucleic acid of interest. Since the ceDNA vector of the invention does not induce an immune response, such a multiple administration strategy will not be impaired by the host immune system response against the ceDNA vector of the invention, contrary to what is observed with encapsidated vectors.

[00380] The ceDNA vector nucleic acid(s) are administered in sufficient amounts to transfect the cells of a desired tissue and to provide sufficient levels of gene transfer and expression without undue adverse effects. Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, intravenous (*e.g.*, in a liposome formulation), direct delivery to the selected organ (*e.g.*, intraportal delivery to the liver), intramuscular, and other parental routes of administration. Routes of administration may be combined, if desired.

[00381] CeDNA vector delivery is not limited to one species of ceDNA vector. As such, in another aspect, multiple ceDNA vectors comprising different exogenous DNA sequences can be delivered simultaneously or sequentially to the target cell, tissue, organ, or subject. Therefore, this strategy can allow for the expression of multiple genes. Delivery can also be performed multiple times and, importantly for gene therapy in the clinical setting, in subsequent increasing or decreasing doses, given the lack of an anti-capsid host immune response due to the absence of a viral capsid. It is anticipated that no anti-capsid response will occur as there is no capsid.

[00382] The invention also provides for a method of treating a disease in a subject comprising introducing into a target cell in need thereof (in particular a muscle cell or tissue) of the subject a therapeutically

effective amount of a ceDNA vector, optionally with a pharmaceutically acceptable carrier. While the ceDNA vector can be introduced in the presence of a carrier, such a carrier is not required. The ceDNA vector implemented comprises a nucleotide sequence of interest useful for treating the disease. In particular, the ceDNA vector may comprise a desired exogenous DNA sequence operably linked to control elements capable of directing transcription of the desired polypeptide, protein, or oligonucleotide encoded by the exogenous DNA sequence when introduced into the subject. The ceDNA vector can be administered via any suitable route as provided above, and elsewhere herein.

XI. Methods of Treatment

[00383] The technology described herein also demonstrates methods for making, as well as methods of using the disclosed ceDNA vectors in a variety of ways, including, for example, ex situ, *in vitro and in vivo* applications, methodologies, diagnostic procedures, and/or gene therapy regimens.

[00384] Provided herein is a method of treating a disease or disorder in a subject comprising introducing into a target cell in need thereof (for example, a muscle cell or tissue, or other affected cell type) of the subject a therapeutically effective amount of a ceDNA vector, optionally with a pharmaceutically acceptable carrier. While the ceDNA vector can be introduced in the presence of a carrier, such a carrier is not required. The ceDNA vector implemented comprises a nucleotide sequence of interest useful for treating the disease. In particular, the ceDNA vector may comprise a desired exogenous DNA sequence operably linked to control elements capable of directing transcription of the desired polypeptide, protein, or oligonucleotide encoded by the exogenous DNA sequence when introduced into the subject. The ceDNA vector can be administered via any suitable route as provided above, and elsewhere herein.

[00385] Any transgene, may be delivered by the ceDNA vectors as disclosed herein. Transgenes of interest include nucleic acids encoding polypeptides, or non-coding nucleic acids (*e.g.*, RNAi, miRs etc.) preferably therapeutic (*e.g.*, for medical, diagnostic, or veterinary uses) or immunogenic (*e.g.*, for vaccines) polypeptides.

[00386] In certain embodiments, the transgenes to be expressed by the ceDNA vectors described herein will express or encode one or more polypeptides, peptides, ribozymes, peptide nucleic acids, siRNAs, RNAis, antisense oligonucleotides, antisense polynucleotides, antibodies, antigen binding fragments, or any combination thereof.

[00387] In particular, the transgene can encode one or more therapeutic agent(s), including, but not limited to, for example, protein(s), polypeptide(s), peptide(s), enzyme(s), antibodies, antigen binding fragments, as well as variants, and/or active fragments thereof, agonists, antagonists, mimetics for use in the treatment, prophylaxis, and/or amelioration of one or more symptoms of a disease, dysfunction, injury, and/or disorder. In one aspect, the disease, dysfunction, trauma, injury and/or disorder is a human disease, dysfunction, trauma, injury, and/or disorder.

[00388] As noted herein, the transgene can encode a therapeutic protein or peptide, or therapeutic nucleic acid sequence or therapeutic agent, including but not limited to one or more agonists, antagonists, anti-

apoptosis factors, inhibitors, receptors, cytokines, cytotoxins, erythropoietic agents, glycoproteins, growth factors, growth factor receptors, hormones, hormone receptors, interferons, interleukins, interleukin receptors, nerve growth factors, neuroactive peptides, neuroactive peptide receptors, proteases, protease inhibitors, protein decarboxylases, protein kinases, protein kinase inhibitors, enzymes, receptor binding proteins, transport proteins or one or more inhibitors thereof, serotonin receptors, or one or more uptake inhibitors thereof, serpins, serpin receptors, tumor suppressors, diagnostic molecules, chemotherapeutic agents, cytotoxins, or any combination thereof.

[00389] In some embodiments, a transgene in the expression cassette, expression construct, or ceDNA vector described herein can be codon optimized for the host cell. As used herein, the term "codon optimized" or "codon optimization" refers to the process of modifying a nucleic acid sequence for enhanced expression in the cells of the vertebrate of interest, *e.g.,* mouse or human (*e.g.*, humanized), by replacing at least one, more than one, or a significant number of codons of the native sequence (*e.g.,* a prokaryotic sequence) with codons that are more frequently or most frequently used in the genes of that vertebrate. Various species exhibit particular bias for certain codons of a particular amino acid. Typically, codon optimization does not alter the amino acid sequence of the original translated protein. Optimized codons can be determined using *e.g.,* Aptagen's Gene Forge® codon optimization and custom gene synthesis platform (Aptagen, Inc.) or another publicly available database.

[00390] Disclosed herein are ceDNA vector compositions and formulations that include one or more of the ceDNA vectors of the present invention together with one or more pharmaceutically-acceptable buffers, diluents, or excipients. Such compositions may be included in one or more diagnostic or therapeutic kits, for diagnosing, preventing, treating or ameliorating one or more symptoms of a disease, injury, disorder, trauma or dysfunction. In one aspect the disease, injury, disorder, trauma or dysfunction is a human disease, injury, disorder, trauma or dysfunction.

[00391] Another aspect of the technology described herein provides a method for providing a subject in need thereof with a diagnostically- or therapeutically-effective amount of a ceDNA vector, the method comprising providing to a cell, tissue or organ of a subject in need thereof, an amount of the ceDNA vector as disclosed herein; and for a time effective to enable expression of the transgene from the ceDNA vector thereby providing the subject with a diagnostically- or a therapeutically-effective amount of the protein, peptide, nucleic acid expressed by the ceDNA vector. In a further aspect, the subject is human.

[00392] Another aspect of the technology described herein provides a method for diagnosing, preventing, treating, or ameliorating at least one or more symptoms of a disease, a disorder, a dysfunction, an injury, an abnormal condition, or trauma in a subject. In an overall and general sense, the method includes at least the step of administering to a subject in need thereof one or more of the disclosed ceDNA vectors, in an amount and for a time sufficient to diagnose, prevent, treat or ameliorate the one or more symptoms of the disease, disorder, dysfunction, injury, abnormal condition, or trauma in the subject. In a further aspect, the subject is human.

[00393] Another aspect is use of the ceDNA vector as a tool for treating or reducing one or more symptoms of a disease or disease states. There are a number of inherited diseases in which defective genes are known, and typically fall into two classes: deficiency states, usually of enzymes, which are generally inherited in a recessive manner, and unbalanced states, which may involve regulatory or structural proteins, and which are typically but not always inherited in a dominant manner. For deficiency state diseases, ceDNA vectors can be used to deliver transgenes to bring a normal gene into affected tissues for replacement therapy, as well, in some embodiments, to create animal models for the disease using antisense mutations. For unbalanced disease states, ceDNA vectors can be used to create a disease state in a model system, which could then be used in efforts to counteract the disease state. Thus the ceDNA vectors and methods disclosed herein permit the treatment of genetic diseases. As used herein, a disease state is treated by partially or wholly remedying the deficiency or imbalance that causes the disease or makes it more severe.

[00394] As still a further aspect, a ceDNA vector as disclosed herein may be employed to deliver a heterologous nucleotide sequence in situations in which it is desirable to regulate the level of transgene expression (*e.g.*, transgenes encoding hormones or growth factors, as described herein).

[00395] Accordingly, in some embodiments, the ceDNA vector described herein can be used to correct an abnormal level and/or function of a gene product (*e.g.*, an absence of, or a defect in, a protein) that results in the disease or disorder. The ceDNA vector can produce a functional protein and/or modify levels of the protein to alleviate or reduce symptoms resulting from, or confer benefit to, a particular disease or disorder caused by the absence or a defect in the protein. For example, treatment of OTC deficiency can be achieved by producing functional OTC enzyme; treatment of hemophilia A and B can be achieved by modifying levels of Factor VIII, Factor IX, and Factor X; treatment of PKU can be achieved by modifying levels of phenylalanine hydroxylase enzyme; treatment of Fabry or Gaucher disease can be achieved by producing functional alpha galactosidase or beta glucocerebrosidase, respectively; treatment of MLD or MPSII can be achieved by producing functional arylsulfatase A or iduronate-2-sulfatase, respectively; treatment of cystic fibrosis can be achieved by producing functional cystic fibrosis transmembrane conductance regulator; treatment of glycogen storage disease can be achieved by restoring functional G6Pase enzyme function; and treatment of PFIC can be achieved by producing functional ATP8B1, ABCB11, ABCB4, or TJP2 genes.

[00396] In alternative embodiments, the ceDNA vectors as disclosed herein can be used to provide an antisense nucleic acid to a cell *in vitro* or *in vivo*. For example, where the transgene is a RNAi molecule, expression of the antisense nucleic acid or RNAi in the target cell diminishes expression of a particular protein by the cell. Accordingly, transgenes which are RNAi molecules or antisense nucleic acids may be administered to decrease expression of a particular protein in a subject in need thereof. Antisense nucleic acids may also be administered to cells *in vitro* to regulate cell physiology, *e.g.*, to optimize cell or tissue culture systems.

[00397] In some embodiments, exemplary transgenes encoded by the ceDNA vector include, but are not limited to: X, lysosomal enzymes (*e.g.*, hexosaminidase A, associated with Tay-Sachs disease, or iduronate

sulfatase, associated, with Hunter Syndrome/MPS II), erythropoietin, angiostatin, endostatin, superoxide dismutase, globin, leptin, catalase, tyrosine hydroxylase, as well as cytokines (*e.g.*, a interferon, β-interferon, interferon-γ, interleukin-2, interleukin-4, interleukin 12, granulocyte-macrophage colony stimulating factor, lymphotoxin, and the like), peptide growth factors and hormones (*e.g.*, somatotropin, insulin, insulin-like growth factors 1 and 2, platelet derived growth factor (PDGF), epidermal growth factor (EGF), fibroblast growth factor (FGF), nerve growth factor (NGF), neurotrophic factor-3 and 4, brain-derived neurotrophic factor (BDNF), glial derived growth factor (GDNF), transforming growth factor-α and -β, and the like), receptors (*e.g.*, tumor necrosis factor receptor). In some exemplary embodiments, the transgene encodes a monoclonal antibody specific for one or more desired targets. In some exemplary embodiments, more than one transgene is encoded by the ceDNA vector. In some exemplary embodiments, the transgene encodes a fusion protein comprising two different polypeptides of interest. In some embodiments, the transgene encodes an antibody, including a full-length antibody or antibody fragment, as defined herein. In some embodiments, the antibody is an antigen-binding domain or a immunoglobulin variable domain sequence, as that is defined herein. Other illustrative transgene sequences encode suicide gene products (thymidine kinase, cytosine deaminase, diphtheria toxin, cytochrome P450, deoxycytidine kinase, and tumor necrosis factor), proteins conferring resistance to a drug used in cancer therapy, and tumor suppressor gene products.

[00398] In a representative embodiment, the transgene expressed by the ceDNA vector can be used for the treatment of muscular dystrophy in a subject in need thereof, the method comprising: administering a treatment, amelioration- or prevention-effective amount of ceDNA vector described herein, wherein the ceDNA vector comprises a heterologous nucleic acid encoding dystrophin, a mini-dystrophin, a micro-dystrophin, myostatin propeptide, follistatin, activin type II soluble receptor, IGF-1, anti-inflammatory polypeptides such as the Ikappa B dominant mutant, sarcospan, utrophin, a micro-dystrophin, laminin-α2, α-sarcoglycan, β-sarcoglycan, γ-sarcoglycan, δ-sarcoglycan, IGF-1, an antibody or antibody fragment against myostatin or myostatin propeptide, and/or RNAi against myostatin. In particular embodiments, the ceDNA vector can be administered to skeletal, diaphragm and/or cardiac muscle as described elsewhere herein.

[00399] In some embodiments, the ceDNA vector can be used to deliver a transgene to skeletal, cardiac or diaphragm muscle, for production of a polypeptide (*e.g.*, an enzyme) or functional RNA (*e.g.*, RNAi, microRNA, antisense RNA) that normally circulates in the blood or for systemic delivery to other tissues to treat, ameliorate, and/or prevent a disorder (*e.g.*, a metabolic disorder, such as diabetes (*e.g.*, insulin), hemophilia (*e.g.*, VIII), a mucopolysaccharide disorder (*e.g.*, Sly syndrome, Hurler Syndrome, Scheie Syndrome, Hurler-Scheie Syndrome, Hunter's Syndrome, Sanfilippo Syndrome A, B, C, D, Morquio Syndrome, Maroteaux-Lamy Syndrome, etc.) or a lysosomal storage disorder (such as Gaucher's disease [glucocerebrosidase], Pompe disease [lysosomal acid .alpha.-glucosidase] or Fabry disease [.alpha.-galactosidase A]) or a glycogen storage disorder (such as Pompe disease [lysosomal acid a glucosidase]). Other suitable proteins for treating, ameliorating, and/or preventing metabolic disorders are described above.

[00400] In other embodiments, the ceDNA vector as disclosed herein can be used to deliver a transgene in a method of treating, ameliorating, and/or preventing a metabolic disorder in a subject in need thereof. Illustrative metabolic disorders and transgenes encoding polypeptides are described herein. Optionally, the polypeptide is secreted (*e.g.*, a polypeptide that is a secreted polypeptide in its native state or that has been engineered to be secreted, for example, by operable association with a secretory signal sequence as is known in the art).

[00401] In other embodiments, the ceDNA vector as disclosed herein may be used to treat seizures, *e.g.*, to reduce the onset, incidence or severity of seizures. The efficacy of a therapeutic treatment for seizures can be assessed by behavioral (*e.g.*, shaking, ticks of the eye or mouth) and/or electrographic means (most seizures have signature electrographic abnormalities). Thus, the ceDNA vector as disclosed herein can also be used to treat epilepsy, which is marked by multiple seizures over time. In one representative embodiment, somatostatin (or an active fragment thereof) is administered to the brain using the ceDNA vector as disclosed herein to treat a pituitary tumor. According to this embodiment, the ceDNA vector as disclosed herein encoding somatostatin (or an active fragment thereof) is administered by microinfusion into the pituitary. Likewise, such treatment can be used to treat acromegaly (abnormal growth hormone secretion from the pituitary). The nucleic acid (*e.g.*, GenBank Accession No. J00306) and amino acid (*e.g.*, GenBank Accession No. P01166; contains processed active peptides somatostatin-28 and somatostatin-14) sequences of somatostatins as are known in the art. In particular embodiments, the ceDNA vector can encode a transgene that comprises a secretory signal as described in U.S. Pat. No. 7,071,172.

[00402] Another aspect of the invention relates to the use of a ceDNA vector as described herein to produce antisense RNA, RNAi or other functional RNA (*e.g.*, a ribozyme) for systemic delivery to a subject *in vivo*. Accordingly, in some embodiments, the ceDNA vector can comprise a transgene that encodes an antisense nucleic acid, a ribozyme (*e.g.*, as described in U.S. Pat. No. 5,877,022), RNAs that affect spliceosome-mediated trans-splicing (see, Puttaraju *et al.*, (1999) Nature Biotech. 17:246; U.S. Pat. No. 6,013,487; U.S. Pat. No. 6,083,702), interfering RNAs (RNAi) that mediate gene silencing (see, Sharp *et al.*, (2000) Science 287:2431) or other non-translated RNAs, such as "guide" RNAs (Gorman *et al.*, (1998) Proc. Nat. Acad. Sci. USA 95:4929; U.S. Pat. No. 5,869,248 to Yuan *et al.*), and the like.

[00403] In some embodiments, the ceDNA vector can further also comprise a transgene that encodes a reporter polypeptide (*e.g.*, an enzyme such as Green Fluorescent Protein, or alkaline phosphatase). In some embodiments, a transgene that encodes a reporter protein useful for experimental or diagnostic purposes, is selected from any of: β-lactamase, β -galactosidase (LacZ), alkaline phosphatase, thymidine kinase, green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), luciferase, and others well known in the art. In some aspects, ceDNA vectors comprising a transgene encoding a reporter polypeptide may be used for diagnostic purposes or as markers of the ceDNA vector's activity in the subject to which they are administered.

[00404] In some embodiments, the ceDNA vector can comprise a transgene or a heterologous nucleotide sequence that shares homology with, and recombines with a locus on the host chromosome. This approach may be utilized to correct a genetic defect in the host cell.

XII. Administration

[00405] In particular embodiments, more than one administration (*e.g.*, two, three, four or more administrations) may be employed to achieve the desired level of gene expression over a period of various intervals, *e.g.*, daily, weekly, monthly, yearly, etc.

[00406] Exemplary modes of administration of the ceDNA vector disclosed herein includes oral, rectal, transmucosal, intranasal, inhalation (*e.g.*, via an aerosol), buccal (*e.g.*, sublingual), vaginal, intrathecal, intraocular, transdermal, intraendothelial, in utero (or in ovo), parenteral (*e.g.*, intravenous, subcutaneous, intradermal, intracranial, intramuscular [including administration to skeletal, diaphragm and/or cardiac muscle], intrapleural, intracerebral, and intraarticular), topical (*e.g.*, to both skin and mucosal surfaces, including airway surfaces, and transdermal administration), intralymphatic, and the like, as well as direct tissue or organ injection (*e.g.*, to liver, eye, skeletal muscle, cardiac muscle, diaphragm muscle or brain).

[00407] Administration of the ceDNA vector can be to any site in a subject, including, without limitation, a site selected from the group consisting of the brain, a skeletal muscle, a smooth muscle, the heart, the diaphragm, the airway epithelium, the liver, the kidney, the spleen, the pancreas, the skin, and the eye. Administration of the ceDNA vector can also be to a tumor (*e.g.*, in or near a tumor or a lymph node). The most suitable route in any given case will depend on the nature and severity of the condition being treated, ameliorated, and/or prevented and on the nature of the particular ceDNA vector that is being used. Additionally, ceDNA permits one to administer more than one transgene in a single vector, or multiple ceDNA vectors (*e.g.* a ceDNA cocktail).

[00408] A. Dose ranges

[00409] *In vivo* and/or *in vitro* assays can optionally be employed to help identify optimal dosage ranges for use. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the condition, and should be decided according to the judgment of the person of ordinary skill in the art and each subject's circumstances. Effective doses can be extrapolated from dose-response curves derived from in vitro or animal model test systems.

[00410] A ceDNA vector is administered in sufficient amounts to transfect the cells of a desired tissue and to provide sufficient levels of gene transfer and expression without undue adverse effects. Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, those described above in the "Administration" section, such as direct delivery to the selected organ (*e.g.*, intraportal delivery to the liver), oral, inhalation (including intranasal and intratracheal delivery), intraocular, intravenous, intramuscular, subcutaneous, intradermal, intratumoral, and other parental routes of administration. Routes of administration can be combined, if desired.

[00411] The dose of the amount of a ceDNA vector required to achieve a particular "therapeutic effect," will vary based on several factors including, but not limited to: the route of nucleic acid administration, the level of gene or RNA expression required to achieve a therapeutic effect, the specific disease or disorder being treated, and the stability of the gene(s), RNA product(s), or resulting expressed protein(s). One of skill in the art can readily determine a ceDNA vector dose range to treat a patient having a particular disease or disorder based on the aforementioned factors, as well as other factors that are well known in the art.

[00412] Dosage regime can be adjusted to provide the optimum therapeutic response. For example, the oligonucleotide can be repeatedly administered, *e.g.*, several doses can be administered daily or the dose can be proportionally reduced as indicated by the exigencies of the therapeutic situation. One of ordinary skill in the art will readily be able to determine appropriate doses and schedules of administration of the subject oligonucleotides, whether the oligonucleotides are to be administered to cells or to subjects.

[00413] A "therapeutically effective dose" will fall in a relatively broad range that can be determined through clinical trials and will depend on the particular application (neural cells will require very small amounts, while systemic injection would require large amounts). For example, for direct in vivo injection into skeletal or cardiac muscle of a human subject, a therapeutically effective dose will be on the order of from about 1 μg to 100 g of the ceDNA vector. If exosomes or microparticles are used to deliver the ceDNA vector, then a therapeutically effective dose can be determined experimentally, but is expected to deliver from 1 μg to about 100 g of vector.

[00414] Formulation of pharmaceutically-acceptable excipients and carrier solutions is well-known to those of skill in the art, as is the development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens.

[00415] For in vitro transfection, an effective amount of a ceDNA vector to be delivered to cells ($1x10^6$ cells) will be on the order of 0.1 to 100 μg ceDNA vector, preferably 1 to 20 μg, and more preferably 1 to 15 μg or 8 to 10 μg. Larger ceDNA vectors will require higher doses. If exosomes or microparticles are used, an effective in vitro dose can be determined experimentally but would be intended to deliver generally the same amount of the ceDNA vector.

[00416] Treatment can involve administration of a single dose or multiple doses. In some embodiments, more than one dose can be administered to a subject; in fact multiple doses can be administered as needed, because the ceDNA vector elicits does not elicit an anti-capsid host immune response due to the absence of a viral capsid. As such, one of skill in the art can readily determine an appropriate number of doses. The number of doses administered can, for example, be on the order of 1-100, preferably 2-20 doses.

[00417] Without wishing to be bound by any particular theory, the lack of typical anti-viral immune response elicited by administration of a ceDNA vector as described by the disclosure (*i.e.*, the absence of capsid components) allows the ceDNA vector to be administered to a host on multiple occasions. In some embodiments, the number of occasions in which a heterologous nucleic acid is delivered to a subject is in a

range of 2 to 10 times (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, or 10 times). In some embodiments, a ceDNA vector is delivered to a subject more than 10 times.

[00418] In some embodiments, a dose of a ceDNA vector is administered to a subject no more than once per calendar day (*e.g.*, a 24-hour period). In some embodiments, a dose of a ceDNA vector is administered to a subject no more than once per 2, 3, 4, 5, 6, or 7 calendar days. In some embodiments, a dose of a ceDNA vector is administered to a subject no more than once per calendar week (*e.g.*, 7 calendar days). In some embodiments, a dose of a ceDNA vector is administered to a subject no more than bi-weekly (*e.g.*, once in a two calendar week period). In some embodiments, a dose of a ceDNA vector is administered to a subject no more than once per calendar month (*e.g.*, once in 30 calendar days). In some embodiments, a dose of a ceDNA vector is administered to a subject no more than once per six calendar months. In some embodiments, a dose of a ceDNA vector is administered to a subject no more than once per calendar year (*e.g.*, 365 days or 366 days in a leap year).

[00419] B. Unit dosage forms

[00420] In some embodiments, the pharmaceutical compositions can conveniently be presented in unit dosage form. A unit dosage form will typically be adapted to one or more specific routes of administration of the pharmaceutical composition. In some embodiments, the unit dosage form is adapted for administration by inhalation. In some embodiments, the unit dosage form is adapted for administration by a vaporizer. In some embodiments, the unit dosage form is adapted for administration by a nebulizer. In some embodiments, the unit dosage form is adapted for administration by an aerosolizer. In some embodiments, the unit dosage form is adapted for oral administration, for buccal administration, or for sublingual administration. In some embodiments, the unit dosage form is adapted for intravenous, intramuscular, or subcutaneous administration. In some embodiments, the unit dosage form is adapted for intrathecal or intracerebroventricular administration. In some embodiments, the pharmaceutical composition is formulated for topical administration. The amount of active ingredient which can be combined with a carrier material to produce a single dosage form will generally be that amount of the compound which produces a therapeutic effect.

XIII. Various applications

[00421] The compositions and ceDNA vectors provided herein can be used to deliver a transgene for various purposes as described above. In some embodiments, the transgene encodes a protein or functional RNA that is intended to be used for research purposes, *e.g.*, to create a somatic transgenic animal model harboring the transgene, *e.g.*, to study the function of the transgene product. In another example, the transgene encodes a protein or functional RNA that is intended to be used to create an animal model of disease.

[00422] In some embodiments, the transgene encodes one or more peptides, polypeptides, or proteins, which are useful for the treatment, amelioration, or prevention of disease states in a mammalian subject.

The transgene can be transferred (*e.g.*, expressed in) to a patient in a sufficient amount to treat a disease associated with reduced expression, lack of expression or dysfunction of the gene.

[00423] In some embodiments, the ceDNA vectors are envisioned for use in diagnostic and screening methods, whereby a transgene is transiently or stably expressed in a cell culture system, or alternatively, a transgenic animal model.

[00424] Another aspect of the technology described herein provides a method of transducing a population of mammalian cells. In an overall and general sense, the method includes at least the step of introducing into one or more cells of the population, a composition that comprises an effective amount of one or more of the ceDNA disclosed herein.

[00425] Additionally, the present invention provides compositions, as well as therapeutic and/or diagnostic kits that include one or more of the disclosed ceDNA vectors or ceDNA compositions, formulated with one or more additional ingredients, or prepared with one or more instructions for their use.

EXAMPLES

[00426] The following examples are provided by way of illustration not limitation.

EXAMPLE 1: Constructing ceDNA Vectors

[00427] Production of the ceDNA vectors using a polynucleotide construct template is described. For example, a polynucleotide construct template used for generating the ceDNA vectors of the present invention can be a ceDNA-plasmid, a ceDNA-Bacmid, and/or a ceDNA-baculovirus. Without being limited to theory, in a permissive host cell, in the presence of *e.g.*, Rep, the polynucleotide construct template having two ITRs and an expression construct, where at least one of the ITRs is modified, replicates to produce ceDNA vectors. ceDNA vector production undergoes two steps: first, excision ("rescue") of template from the template backbone (*e.g.* ceDNA-plasmid, ceDNA-bacmid, ceDNA-bacliovirus genome etc.) via Rep proteins, and second, Rep mediated replication of the excised ceDNA vector.

[00428] An exemplary method to produce ceDNA vectors is from a ceDNA-plasmid as described herein. Referring to **FIG. 1A and 1B**, the polynucleotide construct template of each of the ceDNA-plasmids includes both a left ITR and a right mutated ITR with the following between the ITR sequences: (i) an enhancer/promoter; (ii) a cloning site for a transgene; (iii) a posttranscriptional response element (*e.g.* the woodchuck hepatitis virus posttranscriptional regulatory element (WPRE)); and (iv) a poly-adenylation signal (*e.g.* from bovine growth hormone gene (BGHpA). Unique restriction endonuclease recognition sites (R1-R6) (shown in FIG. 1A and 1B) were also introduced between each component to facilitate the introduction of new genetic components into the specific sites in the construct. R3 (PmeI) GTTTAAAC (SEQ ID NO: 7) and R4 (PacI) TTAATTAA (SEQ ID NO: 542) enzyme sites are engineered into the cloning site to introduce an open reading frame of a transgene. These sequences were cloned into a pFastBac HT B plasmid obtained from ThermoFisher Scientific.

[00429] In brief, a series of ceDNA vectors were obtained from the ceDNA-plasmid constructs shown in Table 12, using the process shown in **FIGS. 4A-4C**. **Table 12** indicates the number of the corresponding

polynucleotide sequence for each component, including sequences active as replication protein site (RPS) (*e.g.* Rep binding site) on either end of a promoter operatively linked to a transgene. The numbers in **Table 12** refer to SEQ ID NOs in this document, corresponding to the sequences of each component.

[00430] Table 12: Exemplary ceDNA constructs.

Plasmid	ITR-L	Promoter	Transgene	ITR -R
Constuct-1	51	3	Luciferase	2
Construct -2	52	3	Luciferase	1
Construct -3	51	4 w/SV40 intr	Luciferase	2
Construct -4	52	4 w/SV40 intr	Luciferase	1
Construct -5	51	5 w/SV40 intr	Luciferase	2
Construct -6	52	5 w/SV40 intr	Luciferase	1
Construct -7	51	6	Luciferase	2
Construct -8	52	6	Luciferase	1

[00431] In some embodiments, a construct to make ceDNA vectors comprises a promoter which is a regulatory switch as described herein, *e.g.*, an inducible promoter. Other constructs were used to make ceDNA vectors, *e.g.*, constructs 10, constructs 11, constructs 12 and construct 13 (see, *e.g.*, Table 14A) which comprise a MND or HLCR promoter operatively linked to a luciferase transgene.

[00432] Production of ceDNA-bacmids:

[00433] With reference to FIG. 4A, DH10Bac competent cells (MAX EFFICIENCY® DH10Bac™ Competent Cells, Thermo Fisher) were transformed with either test or control plasmids following a protocol according to the manufacturer's instructions. Recombination between the plasmid and a baculovirus shuttle vector in the DH10Bac cells were induced to generate recombinant ceDNA-bacmids. The recombinant bacmids were selected by screening a positive selection based on blue-white screening in *E. coli* (Φ80dlacZΔM15 marker provides α-complementation of the β-galactosidase gene from the bacmid vector) on a bacterial agar plate containing X-gal and IPTG with antibiotics to select for transformants and maintenance of the bacmid and transposase plasmids. White colonies caused by transposition that disrupts the β-galactoside indicator gene were picked and cultured in 10 ml of media.

[00434] The recombinant ceDNA-bacmids were isolated from the *E. coli* and transfected into Sf9 or Sf21 insect cells using FugeneHD to produce infectious baculovirus. The adherent Sf9 or Sf21 insect cells were cultured in 50 ml of media in T25 flasks at 25°C. Four days later, culture medium (containing the P0 virus) was removed from the cells, filtered through a 0.45 μm filter, separating the infectious baculovirus particles from cells or cell debris.

[00435] Optionally, the first generation of the baculovirus (P0) was amplified by infecting naïve Sf9 or Sf21 insect cells in 50 to 500 ml of media. Cells were maintained in suspension cultures in an orbital shaker incubator at 130 rpm at 25 °C, monitoring cell diameter and viability, until cells reach a diameter of 18-19 nm (from a naïve diameter of 14-15 nm), and a density of ~4.0E+6 cells/mL. Between 3 and 8 days post-infection, the P1 baculovirus particles in the medium were collected following centrifugation to remove cells and debris then filtration through a 0.45 μm filter.

[00436] The ceDNA-baculovirus comprising the test constructs were collected and the infectious activity, or titer, of the baculovirus was determined. Specifically, four x 20 ml Sf9 cell cultures at 2.5E+6 cells/ml were treated with P1 baculovirus at the following dilutions: 1/1000, 1/10,000, 1/50,000, 1/100,000, and incubated at 25-27°C. Infectivity was determined by the rate of cell diameter increase and cell cycle arrest, and change in cell viability every day for 4 to 5 days.

[00437] With reference to FIG. 4A, a "Rep-plasmid" that comprises a single Rep protein (*e.g.*, see *e.g.*, FIG. 8A) was produced in a pFASTBAC™-Dual expression vector (ThermoFisher).

[00438] The Rep-plasmid was transformed into the DH10Bac competent cells (MAX EFFICIENCY® DH10Bac™ Competent Cells (Thermo Fisher) following a protocol provided by the manufacturer. Recombination between the Rep-plasmid and a baculovirus shuttle vector in the DH10Bac cells were induced to generate recombinant bacmids ("Rep-bacmids"). The recombinant bacmids were selected by a positive selection that included-blue-white screening in *E. coli* (Φ80dlacZΔM15 marker provides α-complementation of the β-galactosidase gene from the bacmid vector) on a bacterial agar plate containing X-gal and IPTG. Isolated white colonies were picked and inoculated in 10 ml of selection media (kanamycin, gentamicin, tetracycline in LB broth). The recombinant bacmids (Rep-bacmids) were isolated from the *E. coli* and the Rep-bacmids were transfected into Sf9 or Sf21 insect cells to produce infectious baculovirus.

[00439] The Sf9 or Sf21 insect cells were cultured in 50 ml of media for 4 days, and infectious recombinant baculovirus ("Rep-baculovirus") were isolated from the culture. Optionally, the first generation Rep-baculovirus (P0) were amplified by infecting naïve Sf9 or Sf21 insect cells and cultured in 50 to 500 ml of media. Between 3 and 8 days post-infection, the P1 baculovirus particles in the medium were collected either by separating cells by centrifugation or filtration or another fractionation process. The Rep-baculovirus were collected and the infectious activity of the baculovirus was determined. Specifically, four x 20 mL Sf9 cell cultures at $2.5x10^6$ cells/mL were treated with P1 baculovirus at the following dilutions, 1/1000, 1/10,000, 1/50,000, 1/100,000, and incubated. Infectivity was determined by the rate of cell diameter increase and cell cycle arrest, and change in cell viability every day for 4 to 5 days.

[00440] ceDNA vector generation and characterization

[00441] With reference to FIG. 4B, Sf9 insect cell culture media containing either (1) a sample-containing a ceDNA-bacmid or a ceDNA-baculovirus, and (2) Rep-baculovirus described above were then added to a fresh culture of Sf9 cells (2.5E+6 cells/ml, 20ml) at a ratio of 1:1000 and 1:10,000, respectively. The cells were then cultured at 130 rpm at 25°C. 4-5 days after the co-infection, cell diameter and viability are detected. When cell diameters reached 18-20nm with a viability of ~70-80%, the cell cultures were centrifuged, the medium was removed, and the cell pellets were collected. The cell pellets are first resuspended in an adequate volume of aqueous medium, either water or buffer. The ceDNA vector was isolated and purified from the cells using Qiagen MIDI PLUS™ purification protocol (Qiagen, 0.2mg of cell pellet mass processed per column).

[00442] Yields of ceDNA vectors produced and purified from the Sf9 insect cells were initially determined based on UV absorbance at 260nm. Yields of various ceDNA vectors determined based on UV absorbance are provided below in Table 13.

[00443] Table 13: Yield of ceDNA vectors from exemplary constructs.

Construct	Culture Volume	Culture Parameters (Diameter in micrometers)	Yield (mg/L)	Estimated Yield (pg/cell)
construct-1	2x1L	Total: 6.02 x10e6 Viability: 53.3% Diameter: 18.4	15.8	5.23

[00444] ceDNA vectors can be assessed by identified by agarose gel electrophoresis under native or denaturing conditions as illustrated in **FIG. 4D**, where (a) the presence of characteristic bands migrating at twice the size on denaturing gels versus native gels after restriction endonuclease cleavage and gel electrophoretic analysis and (b) the presence of monomer and dimer (2x) bands on denaturing gels for uncleaved material is characteristic of the presence of ceDNA vector.

[00445] Structures of the isolated ceDNA vectors were further analyzed by digesting the DNA obtained from co-infected Sf9 cells (as described herein) with restriction endonucleases selected for a) the presence of only a single cut site within the ceDNA vectors, and b) resulting fragments that were large enough to be seen clearly when fractionated on a 0.8% denaturing agarose gel (>800 bp). As illustrated in **FIG. 4E**, linear DNA vectors with a non-continuous structure and ceDNA vector with the linear and continuous structure can be distinguished by sizes of their reaction products– for example, a DNA vector with a non-continuous structure is expected to produce 1kb and 2kb fragments, while a non-encapsidated vector with the continuous structure is expected to produce 2kb and 4kb fragments.

[00446] Therefore, to demonstrate in a qualitative fashion that isolated ceDNA vectors are covalently closed-ended as is required by definition, the samples were digested with a restriction endonuclease identified in the context of the specific DNA vector sequence as having a single restriction site, preferably resulting in two cleavage products of unequal size (*e.g.*, 1000 bp and 2000 bp). Following digestion and electrophoresis on a denaturing gel (which separates the two complementary DNA strands), a linear, non-covalently closed DNA will resolve at sizes 1000 bp and 2000 bp, while a covalently closed DNA (*i.e.*, a ceDNA vector) will resolve at 2x sizes (2000 bp and 4000 bp), as the two DNA strands are linked and are now unfolded and twice the length (though single stranded). Furthermore, digestion of monomeric, dimeric, and *n*-meric forms of the DNA vectors will all resolve as the same size fragments due to the end-to-end linking of the multimeric DNA vectors (see **FIG. 4D**).

[00447] FIG. 5 provides an exemplary picture of a denaturing gel with ceDNA vectors as follows: construct-1, construct -2, construct-3, construct-4, construct-5, construct-6, construct-7 and construct-8 (all described in Table 12 above), with (+) or without (-) digestion by the endonuclease. Each ceDNA vector from constructs-1 to construct-8 produced two bands (*) after the endonuclease reaction. Their two band

sizes determined based on the size marker are provided on the bottom of the picture. The band sizes confirm that each of the ceDNA vectors produced from plasmids comprising construct-1 to construct-8 has a continuous structure.

[00448] As used herein, the phrase "Assay for the Identification of DNA vectors by agarose gel electrophoresis under native gel and denaturing conditions" refers to an assay to assess the close-endedness of the ceDNA by performing restriction endonuclease digestion followed by electrophoretic assessment of the digest products. One such exemplary assay follows, though one of ordinary skill in the art will appreciate that many art-known variations on this example are possible. The restriction endonuclease is selected to be a single cut enzyme for the ceDNA vector of interest that will generate products of approximately 1/3x and 2/3x of the DNA vector length. This resolves the bands on both native and denaturing gels. Before denaturation, it is important to remove the buffer from the sample. The Qiagen PCR clean-up kit or desalting "spin columns," *e.g.* GE HEALTHCARE ILUSTRA™ MICROSPIN™ G-25 columns are some art-known options for the endonuclease digestion. The assay includes for example, i) digest DNA with appropriate restriction endonuclease(s), 2) apply to *e.g.*, a Qiagen PCR clean-up kit, elute with distilled water, iii) adding 10x denaturing solution (10x = 0.5 M NaOH, 10mM EDTA), add 10X dye, not buffered, and analyzing, together with DNA ladders prepared by adding 10X denaturing solution to 4x, on a 0.8 – 1.0 % gel previously incubated with 1mM EDTA and 200mM NaOH to ensure that the NaOH concentration is uniform in the gel and gel box, and running the gel in the presence of 1x denaturing solution (50 mM NaOH, 1mM EDTA). One of ordinary skill in the art will appreciate what voltage to use to run the electrophoresis based on size and desired timing of results. After electrophoresis, the gels are drained and neutralized in 1x TBE or TAE and transferred to distilled water or 1x TBE/TAE with 1x SYBR Gold. Bands can then be visualized with *e.g.* Thermo Fisher, SYBR® Gold Nucleic Acid Gel Stain (10,000X Concentrate in DMSO) and epifluorescent light (blue) or UV (312nm).

[00449] The purity of the generated ceDNA vector can be assessed using any art-known method. As one exemplary and nonlimiting method, contribution of ceDNA-plasmid to the overall UV absorbance of a sample can be estimated by comparing the fluorescent intensity of ceDNA vector to a standard. For example, if based on UV absorbance 4µg of ceDNA vector was loaded on the gel, and the ceDNA vector fluorescent intensity is equivalent to a 2kb band which is known to be 1µg, then there is 1µg of ceDNA vector, and the ceDNA vector is 25% of the total UV absorbing material. Band intensity on the gel is then plotted against the calculated input that band represents – for example, if the total ceDNA vector is 8kb, and the excised comparative band is 2kb, then the band intensity would be plotted as 25% of the total input, which in this case would be .25µg for 1.0µg input. Using the ceDNA vector plasmid titration to plot a standard curve, a regression line equation is then used to calculate the quantity of the ceDNA vector band, which can then be used to determine the percent of total input represented by the ceDNA vector, or percent purity.

EXAMPLE 2: viral DNA production in ceDNA cells

[00450] ceDNA vectors were also generated from constructs 11, 12, 13 and 14 shown in Table 14A. ceDNA-plasmids comprising constructs 11-14 were generated by molecular cloning methods well known in the art. The plasmids in Table 14A were constructed with the WPRE comprising SEQ ID NO: 8 followed by BGHpA comprising SEQ ID NO: 9 in the 3' untranslated region between the transgene and the right side ITR.

Table 14A

Plasmid	ITR-L	Promoter	Transgene	ITR-R
Construct 11	(SEQ ID NO: 63)	(SEQ ID NO: 70)	Luciferase (SEQ ID NO: 71)	(SEQ ID NO: 1)
Construct 12	(SEQ ID NO: 51)	(SEQ ID NO: 70)	Luciferase (SEQ ID NO: 71)	(SEQ ID NO: 64)
Construct 13	(SEQ ID NO: 63)	(SEQ ID NO: 74)	Luciferase (SEQ ID NO: 71)	(SEQ ID NO: 1)
Construct 14	(SEQ ID NO: 51)	(SEQ ID NO: 74)	Luciferase (SEQ ID NO: 71)	(SEQ ID NO: 64)

[00451] The Backbone vector for constructs for constructs 11-14 is as follows: (i) asymITR –MND – luciferase -wPRE-BGH-polyA-ITR in pFB-HTb (construct 11), (ii) ITR –MND –luciferase –wPRE -BGH-polyA –asymITR in pFB-HTb (construct 12), (iii) asymITR –HLCR-AAT-luc-wPRE(O)-BGH-polyA-ITR in pFB-HTb (construct 13); and ITR –HLCR-AAT-luc-wPRE(O)-BGH-polyA- asymITR in pFB-HTb (construct 14), each construct having at least one asymmetric ITR with respect to each other. These constructs also comprise one or more of the following sequences: wPRE0 (SEQ ID NO:72) and BGH-PolyA sequence (SEQ ID NO:73), or sequences at least 85%, or at least 90% or at least 95% sequence identity thereto.

[00452] Next, ceDNA vector production was performed according to the procedure in FIG. 4A-4C, for example, (a) Generation of recombinant ceDNA-Bacmid DNA and Transfection of insect cell with recombinant ceDNA-Bacmid DNA; (b) generation of P1 stock (low titer), P2 stock (high titer), and determination of virus titer by Quantitative-PCR, to obtain a deliverable of 5 ml, >1E+7 plaque forming or infectious units "pfu" per ml BV Stock, BV Stock COA. ceDNA vector isolation was performed by co-infection of 50 ml insect cells with BV stock for the following pairs of infections: Rep-bacmid as disclosed herein and at least one of the following constructs: construct 11, construct 12, construct 13 and construct 14. ceDNA vector isolation was performed using QIAGEN Plasmid Midi Kit to obtain purified DNA material for further analysis. **Table 14B and Table 14C** show the yield (as detected by OD detection) of ceDNA vector produced from constructs 11-14.

[00453] Table 14B: Yield (as detected by OD detection) of exemplary ceDNA vectors produced from constructs 11-14.

Construct No	**DNA Concentration OD260 and Standard Coefficient 50**	**260/280 ratio**	**total DNA [ug] amount from 50 ml infection (ceDNA production volume)**	**Yield total DNA [mg] per 1 liter (estimate)**
Construct 11	342.7 ng/µl	1.79	8.57	0.171
Construct 12	197.5 ng/µl	1.9	4.54	0.090
Construct 13	145 ng/ul	1.9	3.62	0.072
Construct 14	443.1 ng/ul	1.79	11.08	0.221

[00454] Table 14C shows the amount of DNA material obtained (as detected by OD detection) using the constructs 12 and 14 from Table 14C. The yield of total DNA material was acceptable, compared to typical yields of about 3 mg/L of DNA material from the process in Example 1 (Table 13) above.

Construct #	**A_{230}**	**260/230**	**260/280**	**DNA Conc. OD260 and Standard Coefficient 50**	**Yield ug/0.2g cell pellet**	**Yield total DNA [mg] per 1 liter (estimate)**
14	0.038	2.789	1.860	265 ng/ul	53.0	2.6
12	0.017	6.176	1.842	263 ng/ul	52.6	2.6

EXAMPLE 3: ceDNA vectors express luciferase transgene *in vitro*

[00455] Constructs were generated by introducing an open reading frame encoding the Luciferase reporter gene into the cloning site of ceDNA-plasmid constructs: construct-1, construct -3, construct -5, and construct -7. The ceDNA-plasmids (*see* above in Table 12) including the Luciferase coding sequence are named plasmid construct 1-Luc, c plasmid construct-3-Luc, plasmid construct -5-Luc, and plasmid construct 7-Luc, respectively.

[00456] HEK293 cells were cultured and transfected with 100 ng, 200 ng, or 400 ng of plasmid constructs 1, 3, 5 and 7, using FUGENE® (Promega Corp.) as a transfection agent. Expression of Luciferase from each of the plasmids was determined based on Luciferase activity in each cell culture and the results are provided in **FIG. 6A.** Luciferase activity was not detected from the untreated control cells ("Untreated") or cells treated with Fugene alone ("Fugene"), confirming that the Luciferase activity resulted from gene expression from the plasmids. As illustrated in **FIG. 6A** and **FIG. 6B**, robust expression of Luciferase was detected from constructs 1 and 7. The expression from construct-7 expressed Luciferase with a dose-dependent increase of Luciferase activity being detected.

[00457] Growth and viability of cells transfected with each of the plasmids were also determined and presented in **FIG. 7A** and **FIG. 7B.** Cell growth and viability of transfected cells were not significantly different between different groups of cells treated with different constructs.

[00458] Accordingly, Luciferase activity measured in each group and normalized based on cell growth and viability was not different from Luciferase activity without the normalization. ceDNA-plasmid with construct 1-Luc showed the most robust expression of Luciferase with or without normalization.

[00459] Thus, the data presented in **FIGS. 6A, 6B, 7A and 7B** demonstrate that construct 1, comprising from 5' to 3'– WT-ITR (SEQ ID NO: 51), CAG promoter (SEQ ID NO:3), R3/R4 cloning site (SEQ ID NO:7), WPRE (SEQ ID NO: 8), BGHpA (SEQ ID NO:9) and a modified ITR (SEQ ID NO:2), is effective in producing a ceDNA vector that can express a protein of a transgene within the ceDNA vector.

EXAMPLE 4: *In vivo* protein expression of Luciferase Transgene from ceDNA vectors.

[00460] *In vivo* protein expression of a transgene from ceDNA vectors produced from the constructs 1-8 described above is assessed in mice. The ceDNA vector obtained from ceDNA-plasmid construct 1 (as described in Table 12) was tested and demonstrated sustained and durable luciferase transgene expression in a mouse model following hydrodynamic injection of the ceDNA construct without a liposome, redose (at day 28) and durability (up to Day 42) of exogenous firefly luciferase ceDNA. In different experiments, the luciferase expression of selected ceDNA vectors is assessed *in vivo*, where the ceDNA vectors comprise the luciferase transgene and at least one modified ITR selected from any shown in Tables 10A-10B, or an ITR comprising at least one sequences shown in FIGS. 26A-26B

[00461] *In vivo Luciferase expression:* 5-7 week male CD-1 IGS mice (Charles River Laboratories) are administered 0.35 mg/kg of ceDNA vector expressing luciferase in 1.2 mL volume via i.v. hydrodynamic administration to the tail vein on Day 0. . Luciferase expression is assessed by IVIS imaging on Day 3, 4, 7, 14, 21, 28, 31, 35, and 42. Briefly, mice are injected intraperitoneally with 150 mg/kg of luciferin substrate and then whole body luminescence was assessed via IVIS® imaging.

[00462] IVIS imaging is performed on Day 3, Day 4, Day 7, Day 14, Day 21, Day 28, Day 31, Day 35, and Day 42, and collected organs are imaged *ex vivo* following sacrifice on Day 42.

[00463] During the course of the study, animals are weighed and monitored daily for general health and well-being. At sacrifice, blood is collected from each animal by terminal cardiac stick, and split into two portions and processed to 1) plasma and 2) serum, with plasma snap-frozen and serum used for liver enzyme panel and subsequently snap frozen. Additionally, livers, spleens, kidneys, and inguinal lymph nodes (LNs) are collected and imaged *ex vivo* by IVIS.

[00464] Luciferase expression is assessed in livers by MAXDISCOVERY® Luciferase ELISA assay (BIOO Scientific/PerkinElmer), qPCR for Luciferase of liver samples, histopathology of liver samples and/or a serum liver enzyme panel (VetScanVS2; Abaxis Preventative Care Profile Plus).

EXAMPLE 5: ITR Walk Mutant Screening

[00465] Further analyses of the relationship of ITR structure to ceDNA formation were performed. A series of mutants were constructed to query the impact of specific structural changes on ceDNA formation and ability to express the ceDNA-encoded transgene. Mutant construction, assay of ceDNA formation, and assessment of ceDNA transgene expression in human cell culture are described in further detail below.

A. Mutant ITR Construction

[00466] A library of 31 plasmids with unique asymmetric AAV type II ITR mutant cassettes was designed *in silico* and subsequently evaluated in Sf9 insect cells and human embryonic kidney cells (HEK293). Each ITR cassette contained either a luciferase (LUC) or green fluorescent protein (GFP) reporter gene driven by a p10 promoter sequence for expression in insect cells, and a CAG promoter sequence for expression in mammalian cells. Mutations to the ITR sequence were created on either the right or left ITR region. The library contained 15 right-sided (RS) and 16 left-sided (LS) mutants, disclosed in Table 10A and 10B and FIGS 26A and 26B herein.

[00467] Sf9 suspension cultures were maintained in Sf900 III media (Gibco) in vented 200 mL tissue culture flasks. Cultures were passaged every 48 hours and cell counts and growth metrics were measured prior to each passage using a ViCell Counter (Beckman Coulter). Cultures were maintained under shaking conditions (1" orbit, 130 rpm) at 27° C. Adherent cultures of HEK293 cells were maintained in GlutiMax DMEM (Dulbecco's Modified Eagle Medium, Gibco) with 1% fetal bovine serum and .1% PenStrep in 250 mL culture flasks at 37° C with 5% CO_2. Cultures were trypsinized and passaged every 96 hours. A 1:10 dilution of a 90-100% confluent flask was used to seed each passage.

[00468] ceDNA vectors were generated and constructed as described in Example 1 above. In brief, referring to FIG. 4B, Sf9 cells transduced with plasmid constructs were allowed to grow adherently for 24 hours under stationary conditions at 27° C. After 24 hours, transfected Sf9 cells were infected with Rep vector via baculovirus infected insect cells (BIICs). BIICs had been previously assayed to characterize infectivity and were used at a final dilution of 1:2000. BIICs diluted 1:100 in Sf900 insect cell media were added to each previously transfected cell well. Non-Rep vector BIICs were added to a subset of wells as a negative control. Plates were mixed by gentle rocking on a plate rocker for 2 minutes. Cells were then grown for an additional 48 hours at 27° C under stationary conditions. All experimental constructs and controls were assayed in triplicate.

[00469] After 48 hours the 96-well plate was removed to from the incubator, briefly equilibrated to room temperature, and assayed for luciferase expression (OneGlo Luciferase Assay (Promega Corporation)). Total luminescence was measured using a SpectraMax M Series microplate reader. Replicates were averaged. The results are shown in **FIG. 27**. As expected, the three negative controls (media only, mock transfection lacking donor DNA, and sample that was processed in the absence of Rep-containing baculovirus cells) showed no significant luciferase expression. Robust luciferase expression was observed in each of the mutant samples, indicating that for each sample the ceDNA-encoded transgene was successfully transfected and expressed irrespective of the mutation.

B. Assay of ceDNA Formation

[00470] To ensure that the ceDNA generated in the preceding study was of the expected close-ended structure, experiments were performed to produce sufficient amounts of each ceDNA which could subsequently be tested for proper structure. Briefly, Sf9 suspension cultures were transfected with DNA

belonging to a single ITR mutant plasmid from the library. Cultures were seeded at $1.25x10^6$ cells/mL in Erlenmeyer culture flasks with limited gas exchange. DNA:lipid transfection complexes were prepared using FuGene transfection reagent according to the manufacturer's instructions. Complex mixes were prepared and incubated in the same manner as previously described for the luciferase plate assay, with increased volumes proportionate to the number of cells being transfected. As with the reporter gene assay, a ratio of 4.5:1 (volume reagent/mass DNA) was used. Mock (transfection reagents only) and untreated growth controls were prepared in parallel with experimental cultures. Following the addition of transfection reagents, cultures were allowed to recover for 10-15 minutes at room temperature with gentle swirling before being transferred to a 27°C shaking incubator. After 24 hours of incubation under shaking conditions, cell counts and growth metrics for all flasks (experimental and control) were measured using a ViCell counter (Beckman Coulter). All flasks (except growth control) were infected with Rep-vector-containing BIICs at a final dilution of 1:5,000. A positive control using the established BIIC dual infection procedure for ceDNA production was also prepared. The dual infection culture was seeded with the number of cells equal to the average viable cell count of all experimental cultures. Dual infection control was infected with Rep and reporter gene BIICs at a final dilution of 1:5,000 for each construct, respectively. After infection, cultures were placed back in the incubator under previously described shaking conditions. Cell counts, growth and viability metrics were measured daily for all flasks for 3 days post infection. T=0 timepoint measurements were taken after newly infected cultures had been allowed to recover for ~2 hours under shaking incubation conditions. After 3 days cells were harvested by centrifugation for 15 minutes. Supernatant was discarded, mass of pellets was recorded, and pellets were frozen -80° C until DNA extraction.

[00471] Putative crude ceDNA was extracted from all flasks (experimental and control) using the Qiagen Plasmid Plus Midi Purification kit (Qiagen) according to manufacturers "high yield" protocol. Eluates were quantified using optical density measurements obtained from a NanoDrop OneC (ThermoFisher). The resulting ceDNA extracts were stored at 4°C.

[00472] The foregoing ceDNA extracts were run on a native agarose (1% agarose, 1x TAE buffer) gel prepared with 1:10,000 dilution of SYBR Safe Gel Stain (ThermoFisher Scientific), alongside the TrackIt 1kb Plus DNA ladder. The gel was subsequently visualized using a Gbox Mini Imager under UV/blue lighting. As previously described, two primary bands are expected in ceDNA samples run on native gels: a ~5,500 bp band representing a monomeric species and a ~11,000 bp band corresponding to a dimeric species. All mutant samples were tested and displayed the expected monomer and dimer bands on native agarose gels. The results for a representative sample of the mutants are shown in **FIG. 28**. Putative crude ITR-mutant ceDNA and control extracts from small scale production were further assayed using a coupled restriction digest and denaturing agarose gel to confirm a double stranded DNA structure diagnostic of ceDNA. Each mutant ceDNA is expected to have a single EcoR1 restriction site, and so, if properly formed, to produce two characteristic fragments upon EcoR1 digestion. High-fidelity restriction endonuclease EcoRI (New England Biolabs) was used to digest putative ceDNA extract according to manufacturer's

instructions. Extracts from mock and growth controls were not assayed because spectrophotometric quantification using NanoDrop (ThermoFisher) as well as native agarose gel analysis had revealed there to be no detectible ceDNA/plasmid like product in the eluates. Digested material was purified using Qiagen PCR Clean-up Kit (Qiagen) according to manufacturer's instructions with the exception that purified digested material was eluted in nuclease free water instead of Qiagen Elution Buffer. An alkaline agarose gel (.8% alkaline agarose) was equilibrated in Equilibration Buffer (1mM EDTA, 200 mM NaOH) overnight at 4° C. 10x Denaturing Solution (50mM NaOH, 1 mM EDTA) was added to the samples of the purified ceDNA digests and corresponding un-digested ceDNA (1 ug total) and samples were heated at 65° C for 10 minutes. 10x loading dye (Bromophenol blue, 50% glycerol) was added to each denatured sample and mixed. The TrackIt 1kb Plus DNA ladder (ThermoFisher Scientific) was also loaded on the gel as a reference. The gel was run for ~18hrs at 4°C and constant voltage (25 V), followed by rinsing with de-ionized H_2O and neutralization in 1xTAE (Tris-acetate, EDTA) buffer, pH 7.6, for 20 minutes with gentle agitation. The gel was then transferred to 1x TAE/ 1x SYBR Gold solution for ~1 hour under gentle agitation. The gel was then visualized using a Gbox Mini Imager (Syngene) under UV/blue lighting. Uncut denatured samples were expected to migrate at ~11,000 bp and the EcoRI treated samples were expected to have two bands, one at ~4,000 bp and one at ~6,000 bp.

[00473] All mutant samples had similar results in this experiment. Two significant bands were visible in each sample lane in the EcoR1-treated samples, migrating on the denaturing gel at the expected sizes, in sharp contrast to the undigested mutant samples, which migrated at the expected ~11,000 bp size. **FIG. 27** shows the results for a representative sample of mutants, where two bands above background are seen for each digested mutant sample, in comparison to the single band visible in the undigested mutant samples. Thus, the mutant samples seemed to correctly form ceDNA.

C. Functional Expression in Human Cell Culture

[00474] To assess the functionality of mutant ITR ceDNA produced by the small-scale production process, HEK293 cells were transfected with some representative mutant ceDNA samples. Actively dividing HEK293 cells were plated in 96-well microtiter plates at $3x10^6$ cells per well (80% confluency) and incubated for 24 hours at previously described conditions for adherent HEK293 cultures. After 24 hours, 200 ng total of crude small-scale ceDNA was transfected using Lipofectamine (Invitrogen, ThermoFisher Scientific). Transfection complexes were prepared according to manufacturer's instructions and a total volume of 10 uL transfection complex was used to transfect previously plated HEK293 cells. All experimental constructs and controls were assayed in triplicate. Transfected cells were incubated at previously described conditions for 72 hours. After 72 hours the 96-well plate was removed to from the incubator and allowed to briefly equilibrate to room temperature. The OneGlo Luciferase Assay was performed. After 10 minutes on the orbital shaker, total luminescence was measured using a SpectraMax M Series microplate reader. Replicates were averaged. The results are shown in **FIG. 30**. Each of the tested mutant samples expressed luciferase in human cell culture, indicating that ceDNA was correctly formed and expressed for each sample in the context of human cells.

[00475] EXAMPLE 6: Constructs with Rep78 or Rep68 alone are capable of producing ceDNA

[00476] AAV replication (Rep) gene encodes four nonstructural, or replication (Rep), proteins from the same open reading frame. Rep78, Rep68, Rep52, and Rep40 are named for their apparent molecular weights as estimated from their mobility in SDS-PAGE (Mendelson *et al.*, 1986. J Virol. 60: 823-832). Rep78/68 are translated from mRNAs that originate from a transcription promoter at map unit 5 (P5). Rep78 and Rep68 serve as viral replication initiator proteins, which recognize cognate binding sites within the viral origin of replication, and nick the origin at the terminal resolution site. The nicking event provides a free 3′-hydroxyl group that primes viral DNA synthesis. In addition to DNA-binding and site-specific endonuclease activities, Rep78 and Rep68 have been shown to possess helicase and ATPase activities. The Rep52/40 proteins are translated from mRNAs that originate from a transcription promoter at map unit 19 (P19). The Rep52 and Rep40 proteins mediate virus assembly. The Rep68 and Rep40 proteins differ from their longer counterparts in that they are translated from spliced mRNAs from the P5 and P19 promoters, respectively. Splicing removes 92 amino acid residues from the carboxyl termini of the Rep78 and Rep52 proteins and replaces them with 9 amino acids located at the C termini of Rep68 and Rep40.

[00477] Experiments were carried out to determine if the presence of Rep78 or Rep68 alone is sufficient for ceDNA formation. A point mutation was added to eliminate Rep 52 translation by p19 promoter (M->G and M->T) to investigate the effect of deletion of Rep52/40 on ceDNA formation as described in Example 1 above. Thus, constructs modified with the Rep52 (e.g., amino acid 225 M->G and M->T) point mutation will only show ceDNA product from Rep78. Two additional constructs were made to determine if Rep68 has any activity in ceDNA formation. The Rep68 Met→ Gly (M225G) and Rep68 Met→Thr (M225T) mutants were constructed to remove the internal translation site and c-terminal intron sequence (92 amino acid residues and replacement with 9 amino acids as described above). An additional mutant with a point mutation in nickase activity domain (Y156F) was made. **FIGS. 32A and B** depict a non-denaturing gels showing the presence of the highly stable DNA vectors and characteristic bands confirming the presence of the highly stable close-ended DNA (ceDNA) vector made with a single Rep protein using methods described herein. In **Fig. 32A**, higher amounts of ceDNA vector is produced using a nucleic acid of modified Rep78 with the modification of Rep78 of Met→ Gly (M225G) (lane 1) or Rep Met→Thr (M225T) (lane 2) as compared to the production using nucleic acid encoding wild-type Rep78 (lane 5) where the nucleic acid expresses both the Rep78 protein and the Rep52 protein. No ceDNA vector was produced with Rep78 binding mutants, comprising modifications of Gly (Y156F) (lane 3) or Thr (Y156F) with the nickase mutation (lane 4). **Fig. 32B** further illustrates that the Rep68 Met→ Gly (M225G) and Rep68 Met→Thr (M225T) mutants also produced ceDNA vector, to levels equal to or greater than amounts of ceDNA vector produced using a nucleic acid of modified Rep78 with the modification of Rep78 of Met→ Gly (M225G) or Rep Met→Thr (M225T) and a deletion of the c-terminal intron.

[00478] Accordingly, these experiments demonstrated the Rep78 alone or Rep68 alone was sufficient for ceDNA formation without the present of Rep52 or Rep40.

REFERENCES

[00479] All references listed and disclosed in the specification and Examples, including patents, patent applications, International patent applications and publications are incorporated herein in their entirety by reference.

[00480] REP Sequences

SEQ ID NO. 558 is the amino acid sequence of Rep 40 from AAV1.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Glu Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 558)

SEQ ID NO. 559 is the amino acid sequence of Rep 40 from AAV2.

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
Arg Leu Ala Arg Gly His Ser Leu (SEQ ID NO: 559)

SEQ ID NO. 560 is the amino acid sequence of Rep 40 from AAV3A.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
Leu Ala Arg Gly Gln Pro Phe (SEQ ID NO: 560)

SEQ ID NO. 561 is the amino acid sequence of Rep 40 from AAV3B.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Gln Cys Thr Ser Leu
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
Leu Ala Arg Gly Gln Pro Phe (SEQ ID NO: 561)

SEQ ID NO. 562 is the amino acid sequence of Rep 40 from AAV4.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Asn
Pro Pro Glu Asp Ile Ser Ser Asn Arg Ile Tyr Arg Ile Leu Glu Met
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe

Glu Leu Thr Lys Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Thr His Glu Phe Tyr Val Arg Lys Gly Gly Ala Arg Lys Arg Pro Ala
Pro Asn Asp Ala Asp Ile Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Pro Val Asp Tyr Ala Asp
Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 562)

SEQ ID NO. 563 is the amino acid sequence of Rep 40 from AAV5.

Met Ala Leu Val Asn Trp Leu Val Glu His Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr
Gly Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Thr Lys
Ile Met Ser Leu Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser
Val Pro Glu Asp Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys
Gln Arg Ser Phe Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Val Val Val Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe
Glu Leu Thr Lys Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln
Glu Val Lys Asp Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val
Thr His Glu Phe Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala
Glu Lys Ser Leu Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr
Lys Ser Leu Glu Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro
Arg Ser Ser Asp Val Thr Val Asp Pro Ala Pro Leu Arg Pro Leu Asn
Trp Asn Ser Leu Val Gly Pro Ser Trp (SEQ ID NO: 563)

SEQ ID NO. 564 is the amino acid sequence of Rep 40 from AAV6.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 564)

SEQ ID NO. 565 is the amino acid sequence of Rep 40 from AAV7.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ser
Leu Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala

Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg Pro Ala
Pro Asp Asp Ala Asp Ile Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 565)

SEQ ID NO. 566 is the amino acid sequence of Rep 40 from AAV8.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ser
Leu Pro Ala Asp Ile Thr Gln Asn Arg Ile Tyr Arg Ile Leu Ala Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 566)

SEQ ID NO. 567 is the consensus amino acid sequence of SEQ ID NOs 558-566.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Xaa Ser
Pro Pro Glu Asp Ile Ser Thr Asn Arg Ile Tyr Arg Ile Leu Ala Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Xaa Pro Ser Thr Ser Asp Ala Glu Ala Pro Val Asp Phe Ala Asp
Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 567)

SEQ ID NO. 568 is the amino acid sequence of Rep 52 from AAV1.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Glu Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Leu Gln Met
Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe Asn Ile
Cys Phe Thr His Gly Thr Arg Asp Cys Ser Glu Cys Phe Pro Gly Val
Ser Glu Ser Gln Pro Val Val Arg Lys Arg Thr Tyr Arg Lys Leu Cys
Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala
Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 568)

SEQ ID NO. 569 is the amino acid sequence of Rep 52 from AAV2.

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys
Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu
Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr
Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp
Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln (SEQ ID NO: 569)

SEQ ID NO. 570 is the amino acid sequence of Rep 52 from AAV3A.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Ile Ser Asn Val Cys
Phe Thr His Gly Gln Arg Asp Cys Gly Glu Cys Phe Pro Gly Met Ser
Glu Ser Gln Pro Val Ser Val Val Lys Lys Lys Thr Tyr Gln Lys Leu
Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser
Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 570)

SEQ ID NO. 571 is the amino acid sequence of Rep 52 from AAV3B.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Gln Cys Thr Ser Leu
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Ile Ser Asn Val Cys
Phe Thr His Gly Gln Arg Asp Cys Gly Glu Cys Phe Pro Gly Met Ser
Glu Ser Gln Pro Val Ser Val Val Lys Lys Lys Thr Tyr Gln Lys Leu
Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser
Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 571)

SEQ ID NO. 572 is the amino acid sequence of Rep 52 from AAV4.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Asn
Pro Pro Glu Asp Ile Ser Ser Asn Arg Ile Tyr Arg Ile Leu Glu Met
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Lys Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Thr His Glu Phe Tyr Val Arg Lys Gly Gly Ala Arg Lys Arg Pro Ala
Pro Asn Asp Ala Asp Ile Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Pro Val Asp Tyr Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Val Asp Ile Cys
Phe Thr His Gly Val Met Asp Cys Ala Glu Cys Phe Pro Val Ser Glu
Ser Gln Pro Val Ser Val Val Arg Lys Arg Thr Tyr Gln Lys Leu Cys
Pro Ile His His Ile Met Gly Arg Ala Pro Glu Val Ala Cys Ser Ala
Cys Glu Leu Ala Asn Val Asp Leu Asp Asp Cys Asp Met Glu Gln (SEQ ID NO: 572)

SEQ ID NO. 573 is the amino acid sequence of Rep 52 from AAV5.

Met Ala Leu Val Asn Trp Leu Val Glu His Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr
Gly Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Thr Lys
Ile Met Ser Leu Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser
Val Pro Glu Asp Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys
Gln Arg Ser Phe Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Val Val Val Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe
Glu Leu Thr Lys Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln
Glu Val Lys Asp Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val
Thr His Glu Phe Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala
Glu Lys Ser Leu Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr
Lys Ser Leu Glu Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro
Arg Ser Ser Asp Val Thr Val Asp Pro Ala Pro Leu Arg Pro Leu Asn
Trp Asn Ser Arg Tyr Asp Cys Lys Cys Asp Tyr His Ala Gln Phe Asp
Asn Ile Ser Asn Lys Cys Asp Glu Cys Glu Tyr Leu Asn Arg Gly Lys
Asn Gly Cys Ile Cys His Asn Val Thr His Cys Gln Ile Cys His Gly
Ile Pro Pro Trp Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp
Asp Ala Asn Lys Glu Gln (SEQ ID NO: 573)

SEQ ID NO. 574 is the amino acid sequence of Rep 52 from AAV6.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Leu Gln Met
Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe Asn Ile
Cys Phe Thr His Gly Thr Arg Asp Cys Ser Glu Cys Phe Pro Gly Val
Ser Glu Ser Gln Pro Val Val Arg Lys Arg Thr Tyr Arg Lys Leu Cys
Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala
Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 574)

SEQ ID NO. 575 is the amino acid sequence of Rep 52 from AAV7.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ser
Leu Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg Pro Ala
Pro Asp Asp Ala Asp Ile Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Ile Gln Met
Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe Asn Ile
Cys Phe Thr His Gly Val Arg Asp Cys Leu Glu Cys Phe Pro Gly Val
Ser Glu Ser Gln Pro Val Val Arg Lys Lys Thr Tyr Arg Lys Leu Cys
Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala
Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 575)

SEQ ID NO. 576 is the amino acid sequence of Rep 52 from AAV8.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ser
Leu Pro Ala Asp Ile Thr Gln Asn Arg Ile Tyr Arg Ile Leu Ala Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val

Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Leu Gln Met
Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe Asn Ile
Cys Phe Thr His Gly Val Arg Asp Cys Ser Glu Cys Phe Pro Gly Val
Ser Glu Ser Gln Pro Val Val Arg Lys Arg Thr Tyr Arg Lys Leu Cys
Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala
Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 576)

SEQ ID NO. 577 is the consensus amino acid sequence of SEQ ID NOs 568-576.

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Xaa Ser
Pro Pro Glu Asp Ile Ser Thr Asn Arg Ile Tyr Arg Ile Leu Ala Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Ala Pro Val Asp Phe Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Xaa Gln Met Leu
Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Xaa Asn Ile Cys
Phe Thr His Gly Xaa Arg Asp Cys Xaa Glu Cys Phe Pro Gly Val Ser
Glu Ser Gln Xaa Val Val Arg Lys Arg Thr Tyr Xaa Lys Leu Cys Xaa
Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala Cys
Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 577)

SEQ ID NO. 578 is the amino acid sequence of Rep 68 from AAV1.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Ser Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Ile Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Asp Lys Leu Val Gln Thr Ile Tyr Arg Gly Ile Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Glu Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 578)

SEQ ID NO. 579 is the amino acid sequence of Rep 68 from AAV2.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu
Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
Arg Leu Ala Arg Gly His Ser Leu (SEQ ID NO: 579)

SEQ ID NO. 580 is the amino acid sequence of Rep 68 from AAV3A.

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
Glu Arg Leu Pro Gly Ile Ser Asn Ser Phe Val Asn Trp Val Ala Glu

Lys Glu Trp Asp Val Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Thr Tyr Phe His Leu His Val Leu Ile Glu
Thr Ile Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Leu
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
Leu Ala Arg Gly Gln Pro Phe (SEQ ID NO: 580)

SEQ ID NO. 581 is the amino acid sequence of Rep 68 from AAV3B.

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asn Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Thr Tyr Phe His Leu His Val Leu Ile Glu
Thr Ile Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Leu
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Gln Cys Thr Ser Leu
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
Leu Ala Arg Gly Gln Pro Phe (SEQ ID NO: 581)

SEQ ID NO. 582 is the amino acid sequence of Rep 68 from AAV4.

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Ser Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Asp Ser Tyr Phe His Leu His Ile Leu Val Glu
Thr Val Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Asn
Pro Pro Glu Asp Ile Ser Ser Asn Arg Ile Tyr Arg Ile Leu Glu Met
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Lys Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Thr His Glu Phe Tyr Val Arg Lys Gly Gly Ala Arg Lys Arg Pro Ala
Pro Asn Asp Ala Asp Ile Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Pro Val Asp Tyr Ala Asp
Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 582)

SEQ ID NO. 583 is the amino acid sequence of Rep 68 from AAV5.

Met Ala Thr Phe Tyr Glu Val Ile Val Arg Val Pro Phe Asp Val Glu
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asp Trp Val Thr Gly
Gln Ile Trp Glu Leu Pro Pro Glu Ser Asp Leu Asn Leu Thr Leu Val
Glu Gln Pro Gln Leu Thr Val Ala Asp Arg Ile Arg Arg Val Phe Leu
Tyr Glu Trp Asn Lys Phe Ser Lys Gln Glu Ser Lys Phe Phe Val Gln
Phe Glu Lys Gly Ser Glu Tyr Phe His Leu His Thr Leu Val Glu Thr
Ser Gly Ile Ser Ser Met Val Leu Gly Arg Tyr Val Ser Gln Ile Arg
Ala Gln Leu Val Lys Val Val Phe Gln Gly Ile Glu Pro Gln Ile Asn
Asp Trp Val Ala Ile Thr Lys Val Lys Lys Gly Gly Ala Asn Lys Val

Val Asp Ser Gly Tyr Ile Pro Ala Tyr Leu Leu Pro Lys Val Gln Pro
Glu Leu Gln Trp Ala Trp Thr Asn Leu Asp Glu Tyr Lys Leu Ala Ala
Leu Asn Leu Glu Glu Arg Lys Arg Leu Val Ala Gln Phe Leu Ala Glu
Ser Ser Gln Arg Ser Gln Glu Ala Ala Ser Gln Arg Glu Phe Ser Ala
Asp Pro Val Ile Lys Ser Lys Thr Ser Gln Lys Tyr Met Ala Leu Val
Asn Trp Leu Val Glu His Gly Ile Thr Ser Glu Lys Gln Trp Ile Gln
Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr Gly Asn Ser Arg
Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Thr Lys Ile Met Ser Leu
Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser Val Pro Glu Asp
Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met Asn Gly Tyr Asp
Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys Gln Arg Ser Phe
Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala Thr Thr Gly Lys
Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro Phe Tyr Gly Cys
Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp Cys Val Asp Lys
Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn Lys Val Val Glu
Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg Val Asp Gln Lys
Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val Ile Val Thr Ser
Asn Thr Asn Met Cys Val Val Val Asp Gly Asn Ser Thr Thr Phe Glu
His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe Glu Leu Thr Lys
Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln Glu Val Lys Asp
Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val Thr His Glu Phe
Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala Glu Lys Ser Leu
Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr Lys Ser Leu Glu
Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro Arg Ser Ser Asp
Val Thr Val Asp Pro Ala Pro Leu Arg Pro Leu Asn Trp Asn Ser Leu
Val Gly Pro Ser Trp (SEQ ID NO: 583)

SEQ ID NO. 584 is the amino acid sequence of Rep 68 from AAV6.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Ile Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Asp Lys Leu Val Gln Thr Ile Tyr Arg Gly Ile Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala His Asp
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val

Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 584)

SEQ ID NO. 585 is the amino acid sequence of Rep 68 from AAV7.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Val Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Glu Lys Leu Val Gln Thr Ile Tyr Arg Gly Val Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ser
Leu Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg Pro Ala
Pro Asp Asp Ala Asp Ile Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 585)

SEQ ID NO. 586 is the amino acid sequence of Rep 68 from AAV8.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Arg Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Val Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Glu Lys Leu Gly Pro Asp His Leu Pro Ala Gly Ser Ser Pro Thr
Leu Pro Asn Trp Phe Ala Val Thr Lys Asp Ala Val Met Ala Pro Ala
Gly Gly Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu
Pro Lys Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu
Tyr Ile Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala
Gln His Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn
Leu Asn Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala
Arg Tyr Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser
Glu Lys Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn
Ala Ala Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala
Gly Lys Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly
Pro Ser Leu Pro Ala Asp Ile Thr Gln Asn Arg Ile Tyr Arg Ile Leu

Ala Leu Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly
Trp Ala Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly
Pro Ala Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala
Val Pro Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe
Asn Asp Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met
Thr Ala Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys
Val Arg Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr
Pro Val Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly
Asn Ser Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe
Lys Phe Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr
Lys Gln Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr
Glu Val Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg
Pro Ala Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro
Ser Val Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp
Phe Ala Asp Leu Ala Arg Gly Gln Pro Leu (SEQ ID NO: 586)

SEQ ID NO. 587 is the amino acid sequence of Rep 78 from AAV1.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Ser Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Ile Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Asp Lys Leu Val Gln Thr Ile Tyr Arg Gly Ile Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Glu Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Leu Gln Met
Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe Asn Ile
Cys Phe Thr His Gly Thr Arg Asp Cys Ser Glu Cys Phe Pro Gly Val
Ser Glu Ser Gln Pro Val Val Arg Lys Arg Thr Tyr Arg Lys Leu Cys
Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala
Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO:
587)

SEQ ID NO. 588 is the amino acid sequence of Rep 78 from AAV2.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu
Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys
Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu
Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr
Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp
Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln (SEQ ID NO: 588)

SEQ ID NO. 589 is the amino acid sequence of Rep 78 from AAV3A.

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
Glu Arg Leu Pro Gly Ile Ser Asn Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Asp Val Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Thr Tyr Phe His Leu His Val Leu Ile Glu
Thr Ile Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Leu
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Glu Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Ile Ser Asn Val Cys
Phe Thr His Gly Gln Arg Asp Cys Gly Glu Cys Phe Pro Gly Met Ser
Glu Ser Gln Pro Val Ser Val Val Lys Lys Lys Thr Tyr Gln Lys Leu
Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser
Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 589)

SEQ ID NO. 590 is the amino acid sequence of Rep 78 from AAV3B.

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asn Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Thr Tyr Phe His Leu His Val Leu Ile Glu
Thr Ile Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Leu
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Ser Asn
Pro Pro Glu Asp Ile Thr Lys Asn Arg Ile Tyr Gln Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Glu Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Gln Cys Thr Ser Leu
Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Ile Ser Asn Val Cys
Phe Thr His Gly Gln Arg Asp Cys Gly Glu Cys Phe Pro Gly Met Ser
Glu Ser Gln Pro Val Ser Val Val Lys Lys Lys Thr Tyr Gln Lys Leu
Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser

Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 590)

SEQ ID NO. 591 is the amino acid sequence of Rep 78 from AAV4.

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Ser Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Asp Ser Tyr Phe His Leu His Ile Leu Val Glu
Thr Val Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Asn
Pro Pro Glu Asp Ile Ser Ser Asn Arg Ile Tyr Arg Ile Leu Glu Met
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Lys Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Thr His Glu Phe Tyr Val Arg Lys Gly Gly Ala Arg Lys Arg Pro Ala
Pro Asn Asp Ala Asp Ile Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Pro Val Asp Tyr Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Val Asp Ile Cys
Phe Thr His Gly Val Met Asp Cys Ala Glu Cys Phe Pro Val Ser Glu
Ser Gln Pro Val Ser Val Val Arg Lys Arg Thr Tyr Gln Lys Leu Cys
Pro Ile His His Ile Met Gly Arg Ala Pro Glu Val Ala Cys Ser Ala
Cys Glu Leu Ala Asn Val Asp Leu Asp Asp Cys Asp Met Glu Gln (SEQ ID NO: 591)

SEQ ID NO. 592 is the amino acid sequence of Rep 78 from AAV5.

Met Ala Thr Phe Tyr Glu Val Ile Val Arg Val Pro Phe Asp Val Glu
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asp Trp Val Thr Gly
Gln Ile Trp Glu Leu Pro Pro Glu Ser Asp Leu Asn Leu Thr Leu Val
Glu Gln Pro Gln Leu Thr Val Ala Asp Arg Ile Arg Arg Val Phe Leu
Tyr Glu Trp Asn Lys Phe Ser Lys Gln Glu Ser Lys Phe Phe Val Gln
Phe Glu Lys Gly Ser Glu Tyr Phe His Leu His Thr Leu Val Glu Thr
Ser Gly Ile Ser Ser Met Val Leu Gly Arg Tyr Val Ser Gln Ile Arg
Ala Gln Leu Val Lys Val Val Phe Gln Gly Ile Glu Pro Gln Ile Asn
Asp Trp Val Ala Ile Thr Lys Val Lys Lys Gly Gly Ala Asn Lys Val
Val Asp Ser Gly Tyr Ile Pro Ala Tyr Leu Leu Pro Lys Val Gln Pro
Glu Leu Gln Trp Ala Trp Thr Asn Leu Asp Glu Tyr Lys Leu Ala Ala
Leu Asn Leu Glu Glu Arg Lys Arg Leu Val Ala Gln Phe Leu Ala Glu
Ser Ser Gln Arg Ser Gln Glu Ala Ala Ser Gln Arg Glu Phe Ser Ala

Asp Pro Val Ile Lys Ser Lys Thr Ser Gln Lys Tyr Met Ala Leu Val
Asn Trp Leu Val Glu His Gly Ile Thr Ser Glu Lys Gln Trp Ile Gln
Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr Gly Asn Ser Arg
Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Thr Lys Ile Met Ser Leu
Thr Lys Ser Ala Val Asp Tyr Leu Val Gly Ser Ser Val Pro Glu Asp
Ile Ser Lys Asn Arg Ile Trp Gln Ile Phe Glu Met Asn Gly Tyr Asp
Pro Ala Tyr Ala Gly Ser Ile Leu Tyr Gly Trp Cys Gln Arg Ser Phe
Asn Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala Thr Thr Gly Lys
Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro Phe Tyr Gly Cys
Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp Cys Val Asp Lys
Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Asn Lys Val Val Glu
Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg Val Asp Gln Lys
Cys Lys Ser Ser Val Gln Ile Asp Ser Thr Pro Val Ile Val Thr Ser
Asn Thr Asn Met Cys Val Val Val Asp Gly Asn Ser Thr Thr Phe Glu
His Gln Gln Pro Leu Glu Asp Arg Met Phe Lys Phe Glu Leu Thr Lys
Arg Leu Pro Pro Asp Phe Gly Lys Ile Thr Lys Gln Glu Val Lys Asp
Phe Phe Ala Trp Ala Lys Val Asn Gln Val Pro Val Thr His Glu Phe
Lys Val Pro Arg Glu Leu Ala Gly Thr Lys Gly Ala Glu Lys Ser Leu
Lys Arg Pro Leu Gly Asp Val Thr Asn Thr Ser Tyr Lys Ser Leu Glu
Lys Arg Ala Arg Leu Ser Phe Val Pro Glu Thr Pro Arg Ser Ser Asp
Val Thr Val Asp Pro Ala Pro Leu Arg Pro Leu Asn Trp Asn Ser Arg
Tyr Asp Cys Lys Cys Asp Tyr His Ala Gln Phe Asp Asn Ile Ser Asn
Lys Cys Asp Glu Cys Glu Tyr Leu Asn Arg Gly Lys Asn Gly Cys Ile
Cys His Asn Val Thr His Cys Gln Ile Cys His Gly Ile Pro Pro Trp
Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp Asp Ala Asn Lys
Glu Gln (SEQ ID NO: 592)

SEQ ID NO. 593 is the amino acid sequence of Rep 78 from AAV6.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Ile Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Asp Lys Leu Val Gln Thr Ile Tyr Arg Gly Ile Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala His Asp
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val

Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Leu Gln Met
Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe Asn Ile
Cys Phe Thr His Gly Thr Arg Asp Cys Ser Glu Cys Phe Pro Gly Val
Ser Glu Ser Gln Pro Val Val Arg Lys Arg Thr Tyr Arg Lys Leu Cys
Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala
Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 593)

SEQ ID NO. 594 is the amino acid sequence of Rep 78 from AAV7.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Val Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Glu Lys Leu Val Gln Thr Ile Tyr Arg Gly Val Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ser
Leu Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg Pro Ala
Pro Asp Asp Ala Asp Ile Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Ile Gln Met
Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe Asn Ile
Cys Phe Thr His Gly Val Arg Asp Cys Leu Glu Cys Phe Pro Gly Val
Ser Glu Ser Gln Pro Val Val Arg Lys Lys Thr Tyr Arg Lys Leu Cys
Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala
Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln (SEQ ID NO: 594)

SEQ ID NO. 595 is the amino acid sequence of Rep 78 from AAV8.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Arg Asn Leu Ile

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Val Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Glu Lys Leu Gly Pro Asp His Leu Pro Ala Gly Ser Ser Pro Thr
Leu Pro Asn Trp Phe Ala Val Thr Lys Asp Ala Val Met Ala Pro Ala
Gly Gly Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu
Pro Lys Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu
Tyr Ile Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala
Gln His Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn
Leu Asn Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala
Arg Tyr Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser
Glu Lys Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn
Ala Ala Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala
Gly Lys Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly
Pro Ser Leu Pro Ala Asp Ile Thr Gln Asn Arg Ile Tyr Arg Ile Leu
Ala Leu Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly
Trp Ala Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly
Pro Ala Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala
Val Pro Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe
Asn Asp Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met
Thr Ala Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys
Val Arg Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr
Pro Val Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly
Asn Ser Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe
Lys Phe Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr
Lys Gln Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr
Glu Val Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg
Pro Ala Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro
Ser Val Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp
Phe Ala Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Leu
Gln Met Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe
Asn Ile Cys Phe Thr His Gly Val Arg Asp Cys Ser Glu Cys Phe Pro
Gly Val Ser Glu Ser Gln Pro Val Val Arg Lys Arg Thr Tyr Arg Lys
Leu Cys Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys
Ser Ala Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu
Gln (SEQ ID NO: 595)

SEQ ID NO. 596 is the consensus amino acid sequence of **Rep78** ofSEQ ID NOs 587-595.

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Arg Asn Leu Ile
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Val Leu Val Glu
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
Arg Glu Lys Leu Val Xaa Xaa Ile Tyr Arg Gly Ile Glu Pro Thr Leu
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Xaa Ser
Pro Pro Glu Asp Ile Ser Thr Asn Arg Ile Tyr Arg Ile Leu Ala Leu
Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
Glu Val Lys Glu Phe Phe Arg Trp Ala Ser Asp His Val Thr Glu Val
Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
Ala Asp Pro Ser Thr Ser Asp Ala Glu Ala Pro Val Asp Phe Ala Asp
Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Xaa Gln Met Leu
Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Xaa Asn Ile Cys
Phe Thr His Gly Xaa Arg Asp Cys Xaa Glu Cys Phe Pro Gly Val Ser
Glu Ser Gln Xaa Val Val Arg Lys Arg Thr Tyr Xaa Lys Leu Cys Xaa
Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala Cys
Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln

CLAIMS

1. A DNA vector obtained from a vector polynucleotide, wherein the vector polynucleotide encodes a heterologous nucleic acid operatively positioned between a first inverted terminal repeat DNA polynucleotide sequence (ITR) and a second ITR, wherein at least one of the first ITR and the second ITR comprises a nucleotide sequence corresponding to an AAV Rep binding sequence to induce replication of the DNA vector in a cell in the presence of a single species of Rep protein, the DNA vector being obtainable from a method comprising the steps of:

a. incubating a population of cells harboring the vector polynucleotide, which is devoid of viral capsid coding sequences, in the presence of a single species of Rep protein having at least DNA binding and DNA nicking functionality, under conditions effective and for a time sufficient to induce production of the DNA vector within the cells, wherein the cells do not comprise viral capsid coding sequences, and wherein no other species of Rep proteins are present; and

b. harvesting and isolating the resultant DNA vector from the cells.

2. The DNA vector of claim 1, wherein the cell is not contacted with a nucleotide sequence encoding a second Rep protein.

3. The DNA vector of claim 1, wherein the single Rep protein further has helicase, ligase, and ATPase functionality.

4. The DNA vector of claim 1, wherein the Rep protein is an AAV Rep protein.

5. The DNA vector of claim 4, wherein the Rep protein is selected from any of: an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, or AAV12 Rep protein.

6. The DNA vector of claim 4, wherein the Rep protein is an AAV2 Rep 68 protein.

7. The DNA vector of claim 4, wherein the Rep protein is an AAV2 Rep 78 protein.

8. The DNA vector of claim 7, wherein the Rep 78 protein is encoded by a mutant Rep78 nucleotide sequence that does not have a functional translational initiation codon for Rep 52.

9. The DNA vector of claim 8, wherein the mutant Rep 78 nucleotide sequence encodes a mutant Rep 78 protein which comprises a mutation at amino acid position 225 of SEQ ID NO: 530.

10. The DNA vector of claim 9, wherein amino acid position 225 of SEQ ID NO: 530 is mutated to a glycine (Gly) or threonine (Thr).

11. The DNA vector of claim 8, wherein the mutant Rep 78 nucleotide sequence comprises a sequence of SEQ ID NO: 530, or comprises a sequence having at least 95% sequence identity to SEQ ID NO: 530 and has at least DNA binding and DNA nicking functionality, and does not express a second Rep protein.

12. The DNA vector of claim 1, wherein the ITR is a parvovirus ITR.

13. The DNA vector of claim 12, wherein the parvovirus is a dependovirus.

14. The DNA vector of claim 1, wherein the DNA vector is a non-viral capsid-free double-stranded DNA vector with covalently closed ends (ceDNA vector).

15. The DNA vector of claim 14, wherein the presence of the ceDNA vector isolated from the cells can be confirmed by digesting DNA isolated from the cells with a restriction enzyme having a single recognition site on the DNA vector, and analyzing the digested DNA material on a non-denaturing gel to confirm the presence of characteristic bands of linear and continuous DNA as compared to linear and non-continuous DNA.

16. A DNA vector obtained from a vector polynucleotide, wherein the vector polynucleotide encodes a heterologous nucleic acid operatively positioned between two different inverted terminal repeat sequences (ITRs), wherein at least one of the ITRs is a functional ITR comprising a functional terminal resolution site and a Rep binding site; the presence of a single species of Rep protein inducing replication of the vector polynucleotide and production of the DNA vector in a cell, the DNA vector being obtainable from a method comprising the steps of:

a. incubating a population of cells harboring the vector polynucleotide, which is devoid of viral capsid coding sequences, in the presence of a single species of Rep protein that has at least DNA binding and DNA nicking functionality under conditions effective and for time sufficient to induce production of the DNA vector within the cells, wherein the cells do not comprise any nucleic acid encoding Rep52 or Rep40 within the cells, wherein no other species of Rep are present in the cell; and

b. harvesting and isolating the DNA vector from the cells.

17. A polynucleotide for generating a DNA vector comprising a nucleotide sequence encoding a single species of Rep protein amino acid sequence that has at least DNA binding and DNA nicking functionality operatively linked to at least one expression control sequence.

18. The polynucleotide of claim 17, wherein the Rep protein has helicase, ligase, and ATPase functionality.

19. The polynucleotide of claim 17, wherein the Rep protein is an AAV Rep protein.

20. The polynucleotide of claim 19, wherein the AAV Rep protein is selected from any of: an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, or AAV12 Rep protein.

21. The polynucleotide of claim 19, wherein the AAV Rep protein is an AAV2 Rep protein.

22. The polynucleotide of claim 19, wherein the AAV Rep protein is an AAV2 Rep 78 protein.

23. The polynucleotide of claim 22, wherein the Rep 78 protein is encoded by a mutant Rep78 nucleotide sequence that does not have a functional initiation codon for Rep 52.

24. The polynucleotide of claim 23, wherein the mutant Rep 78 nucleotide sequence encodes a mutant Rep78 protein which comprises a mutation at amino acid position 225 of SEQ ID NO: 530.

25. The polynucleotide of claim 24, wherein amino acid 225 of SEQ ID NO: 530 is mutated to a glycine (Gly) or threonine (Thr).

26. The polynucleotide of claim 23, wherein the mutant Rep 78 nucleotide sequence comprises a sequence of SEQ ID NO: 530, or comprises a sequence having at least 95% sequence identity to SEQ ID NO: 530 and has at least DNA binding and DNA nicking functionality, and does not express a second Rep protein.

27. The polynucleotide of claim 17, wherein the at least one expression control sequence encodes an IE promoter, a ΔIE promoter, or a CMV promoter.

28. The polynucleotide of claim 17, wherein the DNA vector is a non-viral capsid-free double stranded DNA vector with covalently closed ends (ceDNA vector).

29. The polynucleotide of claim 28, wherein presence of the ceDNA vector isolated from the cells can be confirmed by digesting DNA isolated from the cells with a restriction enzyme having a single recognition site on the DNA vector and analyzing the digested DNA material on a non-denaturing gel to

confirm the presence of characteristic bands of linear and continuous DNA as compared to linear and non-continuous DNA.

30. A method of producing a DNA vector, the method comprising

contacting a cell with:

(1) a nucleotide sequence encoding a single species of AAV Rep protein (Rep78 and/or Rep68) that has at least DNA binding and DNA nicking functionality, linked to at least one expression control sequence, wherein the cell does not express any other species of Rep protein and is not contacted with any other species of Rep protein;

(2) a double-stranded DNA construct comprising:

an expression cassette;

a first ITR on the upstream (5'-end) of the expression cassette; and

a second ITR on the downstream (3'-end) of the expression cassette, and

(3) harvesting the DNA vector.

31. The method of claim 30, wherein the Rep protein is an AAV Rep protein.

32. The method of claim 31, wherein the AAV Rep protein is selected from any of: an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, or AAV12 Rep protein.

33. The method of claim 31, wherein the AAV Rep protein is an AAV2 Rep 68 protein.

34. The method of claim 31, wherein the AAV Rep protein is an AAV2 Rep 78 protein.

35. The method of claim 34, wherein Rep 78 protein is encoded by a mutant Rep78 nucleotide sequence that does not have a functional initiation codon for Rep 52.

36. The method of claim 35, wherein the mutant Rep 78 nucleotide sequence encodes a mutant Rep 78 protein which comprises a mutation at amino acid position 225 of SEQ ID NO: 530.

37. The method of claim 36, wherein amino acid 225 of SEQ ID NO: 530 is mutated to a glycine (Gly) or threonine (Thr).

38. The method of claim 35, wherein the mutant Rep 78 nucleotide sequence comprises a sequence of SEQ ID NO: 530, or comprises a sequence having at least 95% sequence identity to SEQ ID NO: 530 and has at least DNA binding and DNA nicking functionality.

39. The method of claim 30, wherein the at least one expression control sequence encodes an IE promoter, a ΔIE promoter, or a CMV promoter.

40. The method of any one of claims 30-39, wherein the double-stranded DNA construct is a bacmid, plasmid, minicircle, or a linear double-stranded DNA molecule.

41. The method of any one of claims 30-40, wherein the first ITR upstream of the expression cassette is a wild-type ITR.

42. The method of any one of claims 30-41, wherein the first ITR upstream of the expression cassette and the second ITR downstream of the expression cassette are symmetrical or substantially symmetrical, or asymmetrical relative to each other.

43. The method of any one of claims 30-42, wherein the ITR sequences are selected from any of those listed in Tables 2, 4A, 4B and 5 of International Patent Application PCT/US18/65242.

44. The method of claim 41, wherein the wild-type ITR comprises a polynucleotide of SEQ ID NO: 51.

45. The method of any one of claims 30-44, wherein the second ITR downstream of the expression cassette is a modified ITR.

46. The method of claim 45, wherein the modified ITR comprises a polynucleotide of SEQ ID NO: 2.

47. The method of any one of claims 30-40, wherein the first ITR upstream of the expression cassette is a modified ITR.

48. The method of claim 47, wherein the modified ITR comprises a polynucleotide of SEQ ID NO: 52.

49. The method of any one of claims 47-48, wherein the second ITR downstream of the expression cassette is a wild-type ITR.

50. The method of claim 49, wherein the wild-type ITR comprises a polynucleotide of SEQ ID NO: 1.

51. The method of any one of claims 30-50, wherein the ITR is a replication-competent.

52. The method of any one of claims 30-51 wherein the ITR is an AAV ITR.

53. The method of any one of claims 30-52, wherein the expression cassette comprises a cis-regulatory element.

54. The method of claim 53, wherein the cis-regulatory element is selected from the group consisting of a posttranscriptional regulatory element, and a BGH poly-A signal.

55. The method of claim 54, wherein the posttranscriptional regulatory element comprises a WHP posttranscriptional regulatory element (WPRE).

56. The method of any of claims 30-39, wherein the expression cassette further comprises a promoter selected from the group consisting of CAG promoter, AAT promoter, LP1 promoter, and EF1a promoter.

57. The method of any one of claims 30-56, wherein said expression cassette comprises polynucleotides of SEQ ID NO: 3,SEQ ID NO: 7,SEQ ID NO: 8 and SEQ ID NO: 9.

58. The method of any one of claims 30-57, wherein said expression cassette further comprises an exogenous sequence.

59. The method of claim 58, wherein the exogenous sequence comprises at least 2000 nucleotides.

60. The method of claim 58 or claim 59, wherein the exogenous sequence encodes a protein.

61. The method of claim 58, wherein the exogenous sequence encodes a reporter protein, therapeutic protein, an antigen, a gene editing protein, or a cytotoxic protein.

62. The method of any of claims 30-61, wherein the DNA vector has a linear and continuous structure.

63. A DNA vector generated by the method of any of claims 30-62.

64. A pharmaceutical composition comprising the DNA vector of claim 63; and optionally, an excipient.

65. A kit for producing a DNA vector, the kit comprising:

an expression construct comprising at least one restriction site for insertion of at least one heterologous nucleotide sequence, or regulatory switch, or both, the at least one restriction site operatively positioned between asymmetric inverted terminal repeat sequences (asymmetric ITRs), wherein at least one of the asymmetric ITRs comprises a functional terminal resolution site and a Rep binding site; and

a vector comprising a polynucleotide sequence that encodes a single species of Rep protein, wherein the vector is suitable for expressing the single species of Rep protein in an insect cell.

66. The kit of claim 65, which is suitable for producing the DNA vector of claim 63.

67. The kit of claim 65 or claim 66, further comprising a population of insect cells which is devoid of viral capsid coding sequences, that in the presence of a single species of Rep protein can induce production of the ceDNA vector.

68. A cell comprising: a nucleotide sequence encoding a single species of AAV Rep protein (Rep78 and/or Rep68) that has at least DNA binding and DNA nicking functionality, operably linked to at least one expression control sequence, wherein the cell does not express any other parvovirus Rep protein (Rep52 or Rep40) and is not contacted with any other species of Rep protein; and optionally a double-stranded DNA construct comprising an expression cassette; a first ITR on the upstream (5'-end) of the expression cassette; and a second ITR on the downstream (3'-end) of the expression cassette.

69. The cell of claim 68, wherein the cell is an insect cell.

70. The cell of claim 69, wherein the insect cell is selected from the group consisting of Sf9, Sf21, Trichoplusia ni cell, and High Five cell.

71. The cell of claim 70, wherein the insect cell is Sf9 cell.

72. The cell of claim 70, wherein the insect cell is High Five cell.

73. The cell of claim 68, wherein the cell is a mammalian cell.

74. The cell of claim 73, wherein the mammalian cell is selected from the group consisting of HEK293, Huh-7, HeLa, HepG2, HeplA, 911, CHO, COS, MeWo, NIH3T3, A549, HT1080, monocytes, and mature and immature dendritic cells.

75. The cell of claim 74, wherein the mammalian cell is HEK293.

76. The cell of claim 68, wherein the nucleotide sequence encoding a single species of AAV Rep protein encodes Rep78 and/or Rep68.

77. The cell of claim 76, wherein the nucleotide sequence does not have a functional initiation codon for Rep52 or Rep40.

78. The cell of claim 77, wherein the nucleotide sequence encodes Rep78 protein.

79. The cell of claim 77, wherein the nucleotide sequence encodes Rep68 protein.

80. The cell of claim 77, wherein the nucleotide sequence encodes a mutant Rep78 or Rep68 protein which comprises a mutation at amino acid position 225 of SEQ ID NO: 530.

81. The method of claim 80, wherein amino acid 225 (methionine) of SEQ ID NO: 530 is mutated to a glycine (Gly) or threonine (Thr).

82. The method of claim 80, wherein the nucleotide sequence further comprises one or more modifications in alternative splicing sites in the carboxy terminus, preventing a splicing event leading to production of Rep68, thereby enabling production of Rep78 only.

83. The cell of claim 77, wherein the nucleotide sequence is full length and contains intact alternative splicing sites in the carboxy terminal end, resulting in production of both Rep78 and Rep68.

84. The cell of claim 77, wherein the nucleotide sequence containing a deletion of a carboxy terminal intron sequence, resulting in production of Rep68 only.

85. The cell of claim 77, wherein the nucleotide sequence comprises a sequence of SEQ ID NO: 530, or comprises a sequence having at least 95% sequence identity to SEQ ID NO: 530 and has at least DNA binding and DNA nicking functionality.